# Theoretical Handbook of

# Biostatistics

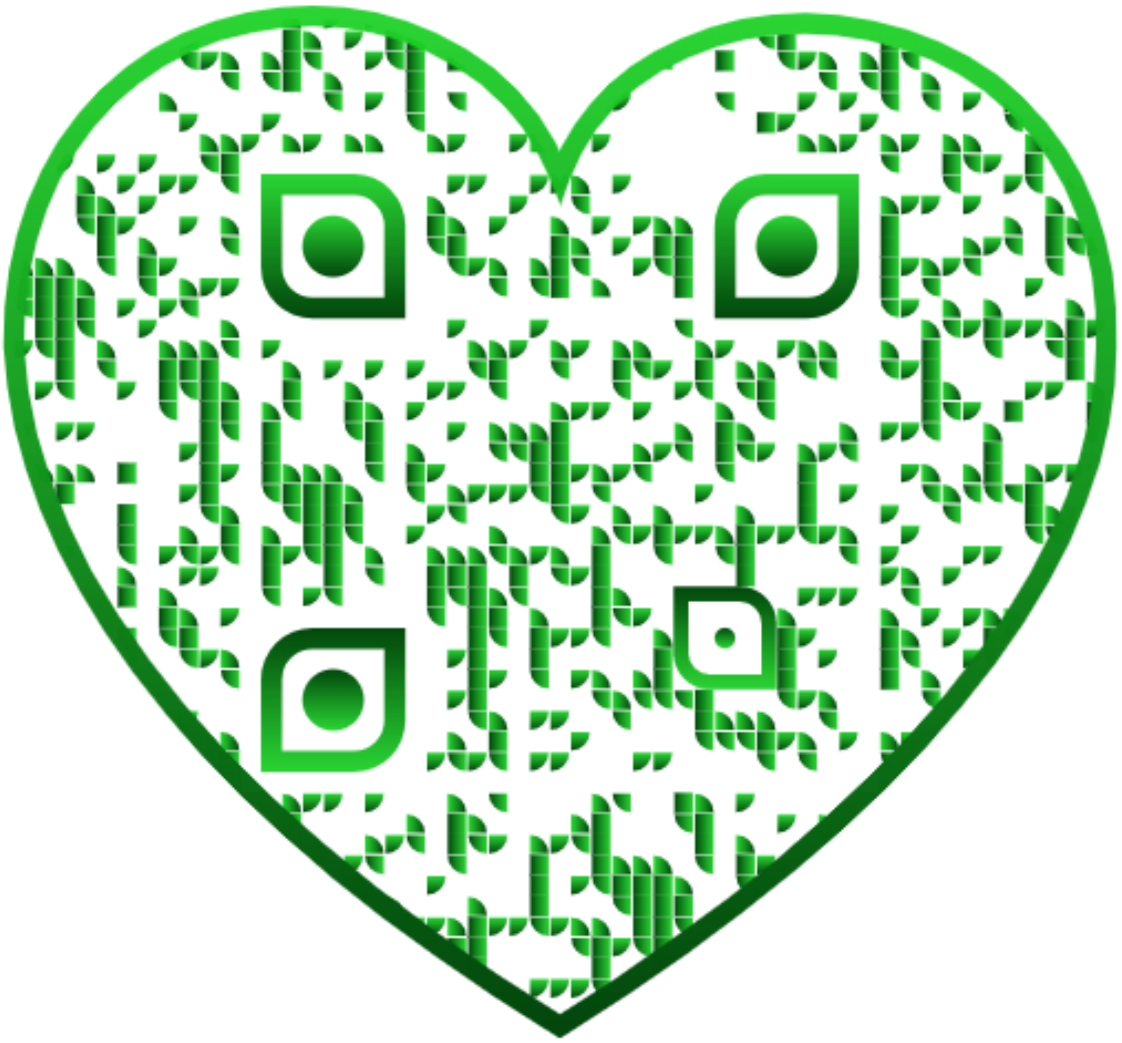
(PH-222)

*Level Two-Second semester*

*By*

*Staff of Pharmacology & Toxicology Department*

*2021/2022*



بكالوريوس الصيدلة (Pharm D– د فارم )

Course Specification

| | |
|---|---|
| **Course name:** Biostatistics | **اسم المقرر** : الاحصاء الحيوى |
| **Academic Level**: Second Level | **المستوى الأكاديمي** : الثانى |
| **Scientific department:** Pharmacology and Toxicology | **القسم العلمي** : الأدوية والسموم |
| **Head of Department:**<br>Prof. Dr. **Ghada M. Suddek** | **رئيس القسم :**<br>ا.د/ غاده محمد صديق |
| **Course Coordinator:**<br>Dr. Hoda Ezzat Kafl | **منسق المقرر :**<br>د/هدى عزت كفل |

| | |
|---|---|
| **University** | Mansoura |
| **Faculty** | Pharmacy |
| **Department offering the course** | Pharmacology and Toxicology |
| **Department supervising the course** | Pharmacology and Toxicology |
| **Program on which the course is given** | Bachelor in Pharmacy- Pharm D |
| **Academic Level** | Level two, second semester, 2021/2022 |
| **Date of course specification approval** | 11th January 2022 |

## A. Basic Information: Course data:

| | |
|---|---|
| **Course Title** | Biostatistics |
| **Course Code** | PH-222 |
| **Prerequisite** | ——— |
| **Teaching credit Hours: Lecture** | 1 |
| **Practical** | ——— |
| **Total Credit Hours** | 1 |

## B. Professional Information:

### 1. Course Aims:

This course enables the students to:

- to make interpretation of any data using statistical analysis,
- determine different methods of sampling, and
- handle the results of different experimental and research studies using suitable statistical techniques.

### 2- Course k. elements:

Upon completing the course, the student will be able to dominate the following key elements

### Domain 1- Fundamental Knowledge

| Program K. element no. | Course K. element no. | Course K. element |
|---|---|---|
| 1.1.1 | 1.1.1.1 | Recognize the important role of biostatistics in biomedical science. |
| 1.1.3 | 1.1.3.1 | Determine different methods of sampling. |
| 1.1.4 | 1.1.4.1 | Show ability to organize data for appropriate statistical analysis and calculation. |

## Domain 2: Professional and Ethical Practice

| Program K. element no. | Course K. element no. | Course K. element |
|---|---|---|
| 2.2.1 | 2.2.1.1 | Differentiate between types of variables. |
| 2.2.3 | 2.2.3.1 | Interpret dispersion of data around the mean. |

## Domain 3: Pharmaceutical Care

| Program K. element no. | Course K. element no. | Course K. element |
|---|---|---|
| 3.3.3 | 3.3.3.1 | Understand basic concepts of probability and statistics. |

## Domain 4: Personal Practice:

| Program K. element no. | Course K. element no. | Course K. element |
|---|---|---|
| **4.1.2** | **4.1.2.1** | Have competence in background mathematics. |

## 3- Course Contents:

| Week No. | Topics | Lecture credit Hours |
|---|---|---|
| **1** | Definition of Biostatistics | 1 |
| **2** | Methods of Data Collection | 1 |
| **3** | Methods of Data Presentation<br>a)Mathematical Presentation | 1 |
| **4** | b) Graphical Presentation<br>c) Tabular Presentation | 1 |
| **5** | Probability and Statistical Hypothesis (Null and Alternative Hypotheses) | 1 |
| **6** | **Periodical (Mid-term exam)** | |
| **7** | Unpaired t test | 1 |
| **8** | Paired t test | 1 |
| **9** | Chi Square | 1 |
| **10** | ANOVA | 1 |
| **11** | Regression (self-learning). | 1 |
| **12** | Revision | |
| **15-16** | **Final written and oral exam** | |

## 4- Teaching and learning Methods:

| | |
|---|---|
| 5.1 | Computer aided learning:<br>a. Lectures using Data show, power Point presentations<br>b. Distance learning<br>• On line learning through my mans "Mansoura university "as recorded – video lectures<br>• Inter active discussion through My Mans |
| 5.2 | Self-learning |
| 5.3 | Class Activity: Group discussion offline and online. |
| 5.4 | Problem – based learning and brainstorming |

## 5- Student Assessment:

### a- Assessment Methods:

| Assessment Methods | K elements to be assessed |
|---|---|
| 1-Written exam | 1.1.1.1, 1.1.3.1, 1.1.4.1, 2.2.1.1, 2.2.3.1, 3.3.3.1 |
| 2-Oral | 1.1.1.1, 1.1.8.1, 2.1.1.1, 2.4.1.1, 4.3.1.1 |
| 3- Periodical (Mid-term exam) / Course work | 1.1.1.1, 1.1.3.1, 1.1.4.1, 2.2.1.1, 2.2.3.1, 3.3.3.1 |

### b. Assessment schedule

| | | |
|---|---|---|
| Assessment 1 | Periodical (Mid-term exam) / Course work | 6th week |
| Assessment 2 | Written exam | 15th -16th week |
| Assessment 3 | Oral exam | 15th -16th week |

### c. Weighing of assessments

| | | |
|---|---|---|
| 1 | Periodical (Mid-term) exam / Course work | 15% |
| 3 | Final-term written examination | 75% |
| 4 | Oral examination | 10% |
| Total | | 100% |

## 6- Facilities required for teaching and learning

| | |
|---|---|
| -Class room | Data show- Computers, Internet. |

## 8- List of References

| No | Reference | Type |
|---|---|---|
| 1. | Electronic book prepared by staff members | Course notes |
| 2. | Recorded videos prepared by stuff members | Videos on platform |
| 3. | Essential Statistics for the Pharmaceutical Sciences, 2nd Edition, Philip Rowe, 2015 Wiley-Blackwell. | Book |
| 4. | Medical Statistics at a Glance, 4th Edition, Aviva Petrie, Caroline Sabin, 2019 Wiley-Blackwell. | Book |
| 5. | http://www.sciencedirect.com / http://www.google scholar.com / http://www.pubmed.com https://www.ekb.eg | websites |

| | |
|---|---|
| **Course Coordinator** | **Dr. Hoda Ezzat Kafl** |
| | |
| **Head of Department** | **Prof. Dr. Ghada M. Suddek** |

**Date: 11 / 1 / 2022**

## 7- Matrix of course content versus course k. elements:

| Week No. | Course contents / K. elements | Domain 1 | | | | Domain 2 | | | Domain 3 | | Domain 4 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | 1.1.1.1 | 1.1.3.1 | 1.1.4.1 | | 2.2.1.1 | 2.2.3.1 | | 3.3.3.1 | | 4.1.2.1 |
| 1 | Definition of Biostatistics | ✓ | | | | | | | | | |
| 2 | Methods of Data Collection | | ✓ | ✓ | | | | | | | ✓ |
| 3 | Methods of Data Presentation<br>a)Mathematical Presentation | | | ✓ | | ✓ | ✓ | | | | ✓ |
| 4 | b) Graphical Presentation<br>c) Tabular Presentation | ✓ | | ✓ | | ✓ | ✓ | | | | |
| 5 | Probability and Statistical Hypothesis (Null and Alternative Hypotheses) | | | ✓ | | ✓ | ✓ | | ✓ | | |
| 7 | Unpaired t test | | ✓ | ✓ | | | ✓ | | ✓ | | |
| 8 | Paired t test | | ✓ | ✓ | | | ✓ | | ✓ | | |
| 9 | Chi Square | | ✓ | ✓ | | | ✓ | | ✓ | | |
| 10 | ANOVA | | ✓ | ✓ | | | ✓ | | ✓ | | |
| 11 | Regression (self-learning). | | ✓ | ✓ | | ✓ | ✓ | | ✓ | | |
| 12 | Revision | ✓ | ✓ | ✓ | | ✓ | ✓ | | ✓ | | ✓ |

# Introduction

**Statistics** is the science of collecting, summarizing, presenting and analyzing data. This leadsto conclusion and subsequent decision. When the data analyzed are derived from the biological science and medicine, we use the term **biostatistics** to distinguish this particular application of statistical tools and concepts.

## Data:

- The raw material of Statistics is data.
- We may define data as figures. Figures result from the process of counting or from taking a measurement.
- *For example:*
- - When a hospital administrator counts the number of patients (counting).
- - When a nurse weighs a patient (measurement)

## Sources of Data:

We search for suitable data to serve as the raw material for our investigation.
Such data are available from one or more of the following sources:

1- Routinely kept records.

- Hospital medical records contain immense amounts of information on patients.
- Hospital accounting records contain a wealth of data on the facility's business

2- External sources.

The data needed to answer a question may already exist in the form of published reports, commercially available data banks, or the research literature, i.e. someone else has already asked the same question.

**3- Surveys:**

The source may be a survey, if the data needed is about answering certain questions.

*For example:*: If the administrator of a clinic wishes to obtain information regarding the mode of transportation used by patients to visit the clinic, then a survey may be conducted among patients to obtain this information.

**4- Experiments.**

Frequently the data needed to answer a question are available only as the result of an experiment.

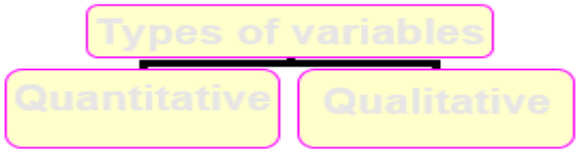
*For example:* If a nurse wishes to know which of several strategies is best for maximizing patient compliance, she might conduct an experiment in which the different strategies of motivating compliance are tried with different patients.

# * A variable:

It is a characteristic that takes on different values in different persons, places, or things.

*For example:*

- heart rate,
- the heights of adult males,
- the weights of preschool children,
- the ages of patients seen in a dental clinic.




## Quantitative Variables

It can be measured in the usual sense.

*For example:*

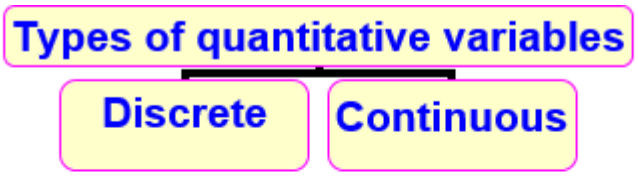
- the heights of adult males,
- the weights of preschool children,
- the ages of patients seen in a dental clinic.

## Qualitative Variables

Many characteristics are not capable of being measured. Some of them can be ordered (called ordinal) and Some of them can't be ordered (called nominal).

*For example:*

- classification of people into socio-economic groups (ordinal),
- .hair color (nominal)




## A discrete variable

is characterized by gaps or interruptions in the values that it can assume.

*For example:*

- The number of daily admissions to a general hospital,
- The number of decayed, missing or filled teeth per child in an elementary school.

## A continuous variable

can assume any value within a specified relevant interval of values assumed by the variable.

*For example:*

- Height,
- weight,
- skull circumference.

No matter how close together the observed heights of two people, we can find another person whose height falls somewhere in between.

## * A population:

It is the largest collection of values of a random variable for which we have an interest at a particular time.

*For example:*

The weights of all the children enrolled in a certain elementary school.

Populations may be finite or infinite.

## * A sample:

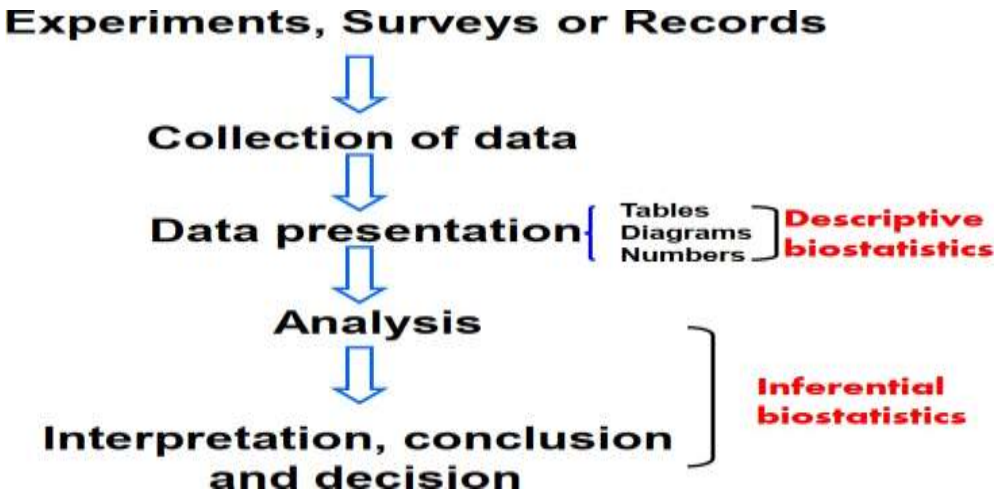
It is a part of a population.

*For example:*

The weights of only a fraction of these children.

**Role of biostatisticians**

- To guide **the design of an experiment** or survey prior to data collection
- To **analyze data** using proper statistical procedures and techniques
- To **present and interpret** the results to researchers and other decision makers



Strategies for understanding the meanings of Data:

*Descriptive Statistics*

## Frequency Distribution for Discrete Random Variables

*Example:*

Suppose that we take a sample of size 16 from children in a primary school and get the following data about the number of their decayed teeth,

3,5,2,4,0,1,3,5,2,3,2,3,3,2,4,1

To construct a frequency table:

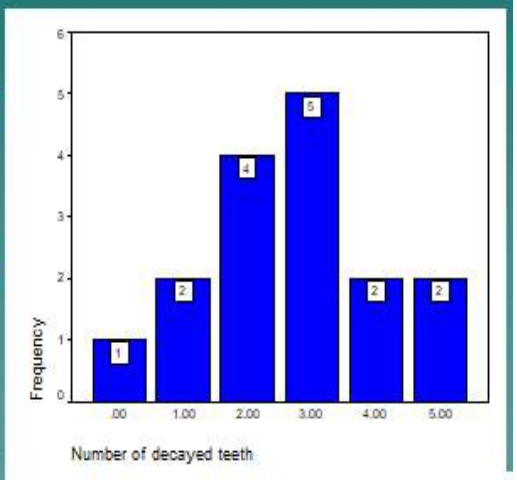
1- Order the values from the smallest to the largest.

0,1,1,2,2,2,2,3,3,3,3,3,4,4,5,5

2- Count how many numbers are the same.

| No. of decayed teeth | Frequency | Relative Frequency |
|---|---|---|
| 0 | 1 | 0.0625 |
| 1 | 2 | 0.125 |
| 2 | 4 | 0.25 |
| 3 | 5 | 0.3125 |
| 4 | 2 | 0.125 |
| 5 | 2 | 0.125 |
| Total | 16 | 1 |

*Representing the simple frequency table using the bar chart*




## Frequency Distribution for Continuous Random Variables

For large samples, we can't use the simple frequency table to represent the da
We need to divide the data into groups or intervals or classes.
So, we need to determine:

*1- The number of intervals (k).*

Too few intervals are not good because information will be lost.
Too many intervals are not helpful to summarize the data.
A commonly followed rule is that $6 \le k \le 15$,
or the following formula may be used, $k = 1 + 3.322 (\log n)$

*2- The range (R).*

It is the difference between the largest and the smallest observation in the data set.

*3- The Width of the interval (w).*

Class intervals generally should be of the same width. Thus, if we want k intervals, then w is chosen such that $w \ge R / k$.

## Example

- We wish to know how many class interval to have in the frequency distribution of the data in Table 1.4.1 Page 9-10 of ages of 189 subjects who Participated in a study on smoking cessation
- Solution :
- Since the number of observations equal 189, then
- k = 1+3.322(log 169)
- = 1 + 3.3222 (2.276) ≅ 9,
- R = 82 – 30 = 52 and
- w = 52 / 9 = 5.778

- It is better to let w = 10, then the intervals
- will be in the form:

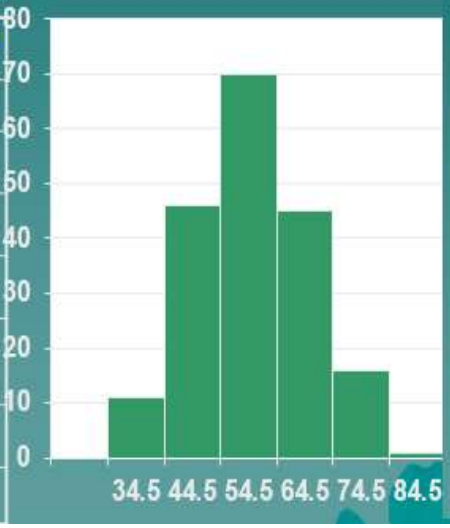
| Class interval | Frequency |
|---|---|
| 30 – 39 | 11 |
| 40 – 49 | 46 |
| 50 – 59 | 70 |
| 60 – 69 | 45 |
| 70 – 79 | 16 |
| 80 – 89 | 1 |
| Total | 189 |

Sum of frequency =sample size=n

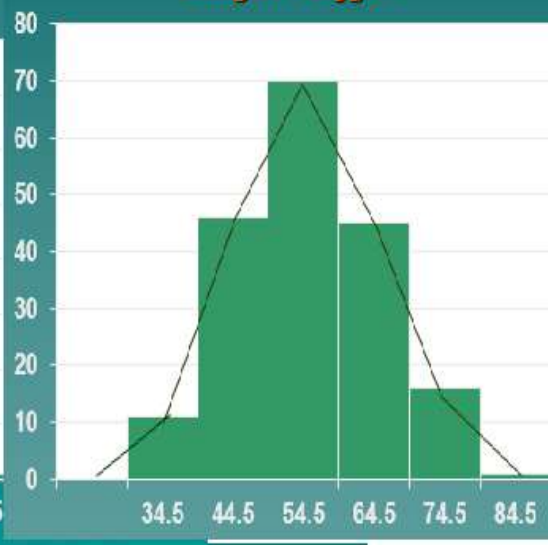
## Representing the grouped frequency table using the histogram

To draw the histogram, the true classes limits should be used. They can be computed by subtracting 0.5 from the lower limit and adding 0.5 to the upper limit for each interval.

| True class limits | Frequency |
|---|---|
| 29.5 – <39.5 | 11 |
| 39.5 – < 49.5 | 46 |
| 49.5 – < 59.5 | 70 |
| 59.5 – < 69.5 | 45 |
| 69.5 – < 79.5 | 16 |
| 79.5 – < 89.5 | 1 |
| Total | 189 |




## Representing the grouped frequency table using the Polygon



## 1- Frequency distribution table (F.D.T.):

**Example 1:** The marks of 52 students in a biostatistics exam were as follows:

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| **2** | **2** | **5** | **1** | **2** | **6** | **3** | **3** | **4** | **2** |
| **4** | **0** | **5** | **7** | **7** | **5** | **6** | **6** | **8** | **10** |
| **7** | **2** | **2** | **10** | **5** | **8** | **2** | **5** | **4** | **2** |
| **6** | **2** | **6** | **1** | **7** | **2** | **7** | **2** | **3** | **8** |
| **1** | **5** | **2** | **5** | **2** | **14** | **2** | **2** | **6** | **3** |
| **1** | | | | | | | | | **7** |

**Constructing A Frequency Table**

1- Arrange data values in an ascending way

0 1 1 1 1 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 3 3 3 3 4 4 4 5 5 5 5 5 5 5 6 6 6 6 6 6 7 7 7 7 7 8 8 8 10 10 14

2- Calculate the range

The range = the highest value - the lowest value = 14 - 0 = 14

3-Decide on the number of the classes (categories) desired

4-Determine the class interval by dividing the range by the number of classes and round up.

| Class | Frequency |
|---|---|
| 0 - | 20 |
| 3 - | 14 |
| 6 - | 15 |
| 9 - | 2 |
| 12 - | 1 |
| | Total = 52 |

**Class interval ≈ round up of** **Range / number of classes** **= 14/5 =2.8 ≈ 3**

| Class | Frequency |
|---|---|
| 0 - | 20 |
| 3 - | 14 |
| 6 - | 15 |
| 9 - | 2 |
| 12 - | 1 |

| Mark | Cumulative Frequency | Percent of the total |
|---|---|---|
| Less than 3 | 20 | 38.5 |
| Less than 6 | 34 | 65 |
| Less than 9 | 49 | 94 |
| Less than 12 | 51 | 98 |
| Less than 15 | 52 | 100 |

**Example 2:**

-In the following table the figures are arranged in an ascending order.

***Table 1:*** *The systolic blood pressure in mm/Hg of 50 healthy individuals.*

| | | | | |
|---|---|---|---|---|
| 110 | 118 | 120 | 122 | 128 |
| 112 | 118 | 120 | 122 | 128 |
| 114 | 119 | 121 | 122 | 128 |
| 115 | 120 | 121 | 123 | 129 |
| 116 | 120 | 121 | 124 | 130 |
| 116 | 120 | 122 | 125 | 134 |
| 118 | 120 | 122 | 126 | 134 |
| 118 | 120 | 122 | 126 | 134 |
| 118 | 120 | 122 | 126 | 135 |
| 118 | 120 | 122 | 126 | 135 |

***Table 2:*** *Frequency distribution table of the systolic blood pressure in mm/Hg of 50 healthy individuals.*

| Class number | Class | Mid-point of the class | Frequency |
|---|---|---|---|
| 1 | 110- | 112.5 | 4 |
| 2 | 115- | 117.5 | 10 |
| 3 | 120- | 122.5 | 21 |
| 4 | 125- | 127.5 | 9 |
| 5 | 130- | 132.5 | 6 |
| | | | Total = 50 |

In this example, as it is evident from table (2), the number of classes is 5, the class interval is 5, the mid-point of the class (1) for example = 110+2.5 = 112.5 mm/Hg. It is also quite clear from this table that the frequency distribution table cangive a summary of the data and good information about it is obtained quickly.

-Thus, from this table we can simply conclude that the most frequent blood pressure of individuals is that lying in the third class having a value of from 120 -125 mm/Hg with a midpoint of 122.5mm/Hg.

### 3- Cumulative frequency distribution tables:

- This type of tables is used when it is required to know how many individuals have a value less or more than a certain value. In the cumulative frequency distribution tables the values are arranged in ascending manner (ascending cumulative frequency tables) or descending manner (descending frequency tables).

***Table 3:*** *Ascending cumulative frequency distribution table of systolic blood pressure(in mm/Hg) of 50 healthy individuals.*

| Systolic blood pressure (mm/Hg) | Number of Individuals | Percent of the total |
|---|---|---|
| Less than 115 | 4 | 8 |
| Less than 120 | 14 | 28 |
| Less than 125 | 35 | 70 |
| Less than 130 | 44 | 88 |
| Less than 135 | 50 | 100 |

***Table 4:*** *Descending cumulative blood pressure (in mm/Hg) of 50 healthy individuals.*

| Systolic blood pressure (mm/Hg) | Number of Individuals | Percent of the total |
|---|---|---|
| More than 130 | 6 | 12 |
| More than 125 | 15 | 30 |
| More than 120 | 36 | 72 |
| More than 115 | 46 | 92 |
| More than 110 | 50 | 100 |
| Remember that in the first type (ascending table, table (3) we start with the second class from table (2). Whereas in the second type (descending type, table (4)we start with the class just before the last (from table 2). | | |

**Descriptive statistics B) Diagrams:**

1. Bars
2. Pie (sector diagram).
3. Histogram.
4. Frequency polygon.
5. Frequency curve.
6. Dot plot
7. Scatter or dot diagram
8. Line graph.
9. Pictogram.

## 1- Bar chart:

It is a graphical presentation of the data by rectangular of constant width, drawn with lengthproportional to frequency, relative frequency or magnitude. Bar charts are used to depict categorical information e.g., Male–Female, Mild–Moderate–Severe, etc.

**a) Simple bar chart:**

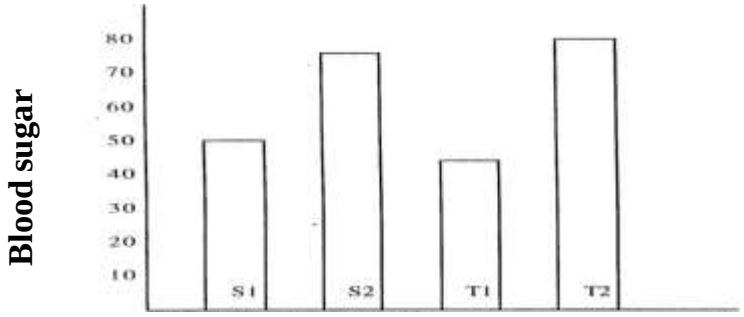
- If we are representing the data of one phenomenon a simple bar chart is used as in the following **example1:** *The blood sugar levels of rabbits after treatment with 2 doses ofthe standard and 2 doses of the test solution of insulin*

-In a biological assay of an unknown solution of insulin on the blood sugar level of rabbits, 4groups of animals were used (S1, S2, T1, T2).

-The following mean blood sugar levels (in mg %) were obtained for each group after treatment with the standard (S) and test (T) solutions of insulin.

| | | | | | |
|---|---|---|---|---|---|
| | Dose of insulin | | | | |
| | Blood sugar (mg %) | | | | |

-Represent these data using bar charts.



**b) Multiple bar chart**:

In an experiment studying the effect of reserpine pretreatment on the concentration of noradrenaline, dopamine and serotonin in the brain, the following results were obtained.

| Amine | Control group | Pretreated group with reserpine |
|---|---|---|
| Noradrenaline | 0.40 | 0.16 |
| Dopamine | 1.06 | 0.50 |
| Serotonin | 0.54 | 0.30 |

- Represent these data using bar charts.

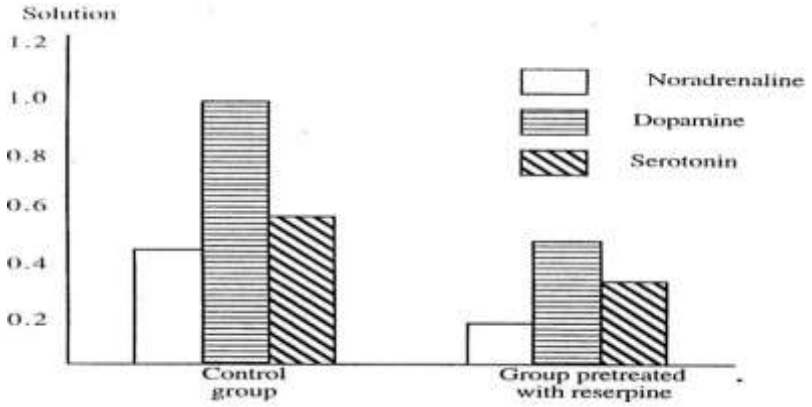



***Figure 2:*** *Multiple bar chart. The effect of pretreatment with reserpine on the concentration ofnoradrenaline, dopamine and serotonin in the brain of rats.*

**c) <u>Component bar chart (Proportional)</u>:**

-It is subdivision of single bar chart to indicate a composition of a total

***Table 5:*** *Monthly distribution of new and repeat patients in a hospital during the year 1973.*

| Months | Number of cases | | |
|---|---|---|---|
| | New | Repeat | Total |
| January | 2658 | 1114 | 3772 |
| February | 2052 | 1470 | 3522 |
| March | 2179 | 1610 | 3789 |
| April | 1980 | 1351 | 3331 |
| May | 1714 | 1175 | 2889 |
| June | 2153 | 1434 | 3587 |
| July | 2203 | 1494 | 3697 |
| August | 2123 | 1536 | 3659 |
| September | 2642 | 1772 | 4414 |
| October | 3055 | 1965 | 5020 |
| November | 3869 | 2360 | 6229 |
| December | 2557 | 1894 | 4451 |

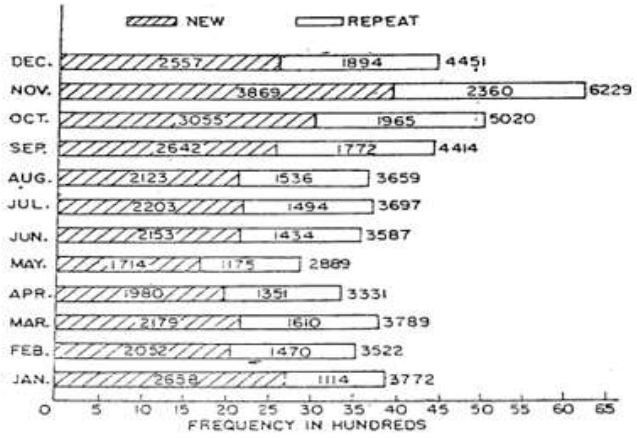



***Figure 3:*** *Proportional bar diagram showing new and repeat outdoor attendance*

**2- Pie or sector diagram:**

This another way of presenting data of qualitative characters such as blood groups, Rhgroups, age groups, sex groups etc….

The frequencies of the groups are shown in a circle. Degrees of angle denote the frequencyand area of the sector.

**Example:** The following is the distribution of blood groups in certain city.

| Blood group | Number |
|---|---|
| A | 744 |
| B | 971 |
| O | 888 |
| A B | 207 |
| Total | 2810 |

-Represent these data by sector diagram.

| Blood group | Number | % | Degree |
|---|---|---|---|
| A | 744 | 26.5 | 95.4 |
| B | 971 | 34.5 | 124.2 |
| O | 888 | 31.6 | 113.8 |
| A B | 207 | 7.4 | 26.6 |
| Total | 2810 | 100.0 | 360.0 |

-Size of angle for blood group A = 26.5 X 3.6 = 95.4 **OR** = (744/1810) X 360 = 95.4

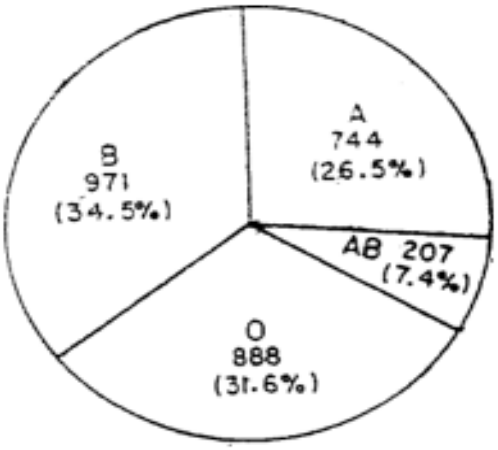



***Figure 4:*** *Pie or sector diagram showing distribution of blood groups*

### 3- Histogram:

- It is a continuous bar chart presenting data in a frequency distribution.

- Distributions are often symmetrical with most scores falling in the middle and fewer toward the extremes

- *Most biological data are symmetrically distributed and form a normal curve (a.k.a, bell-shaped curve)*

The class intervals are in the X-axis, on each class a rectangle is drawn with a corresponding height to the frequency (F) on the Y-axis.

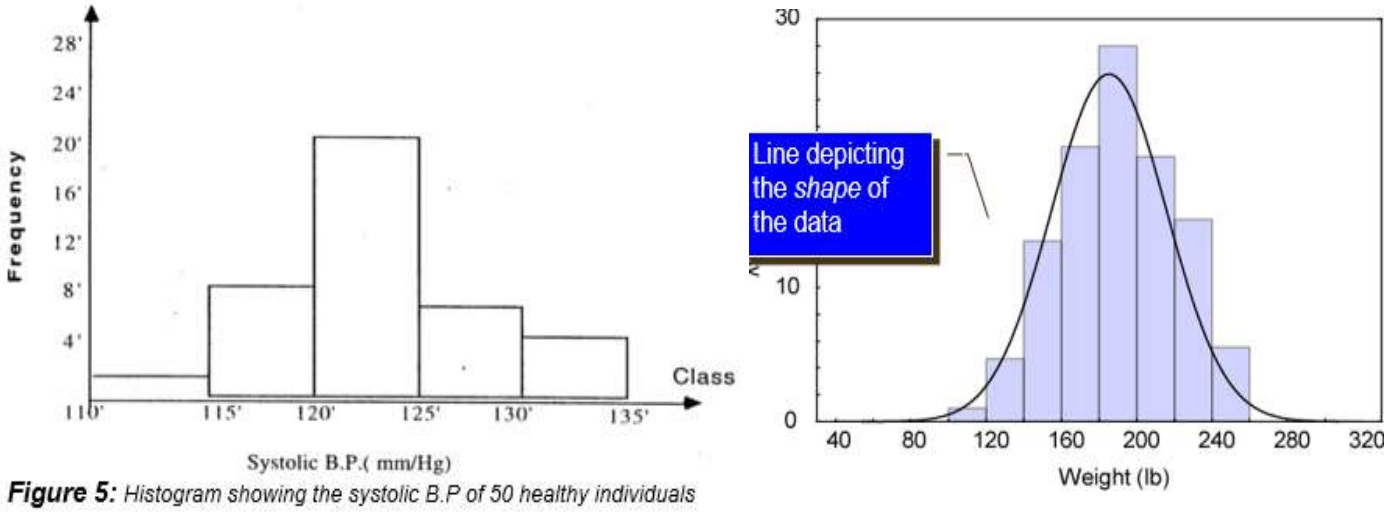



***Figure 5:*** *Histogram showing the systolic B.P of 50 healthy individuals*

### 4- Frequency polygon:

-If in histogram a dot is put in the middle of the upper base of each rectangle and these dots are connected by a straight line, the frequency polygon is obtained. Each point in this line represents the midpoint of the class and its frequency (Figure6).

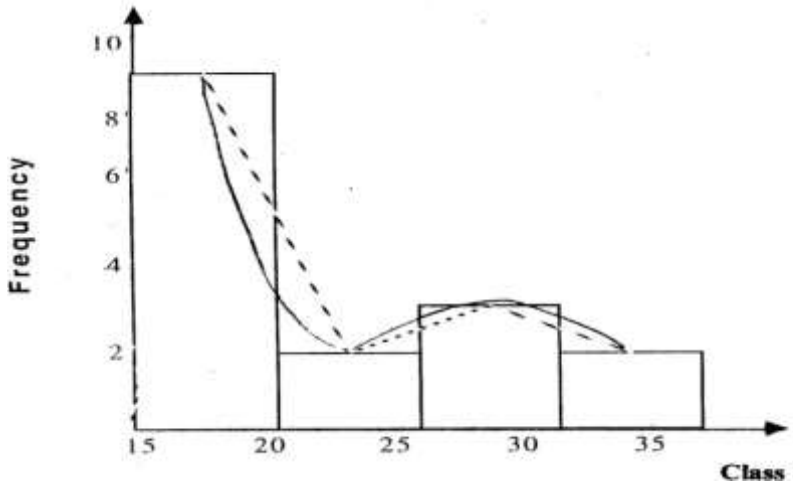
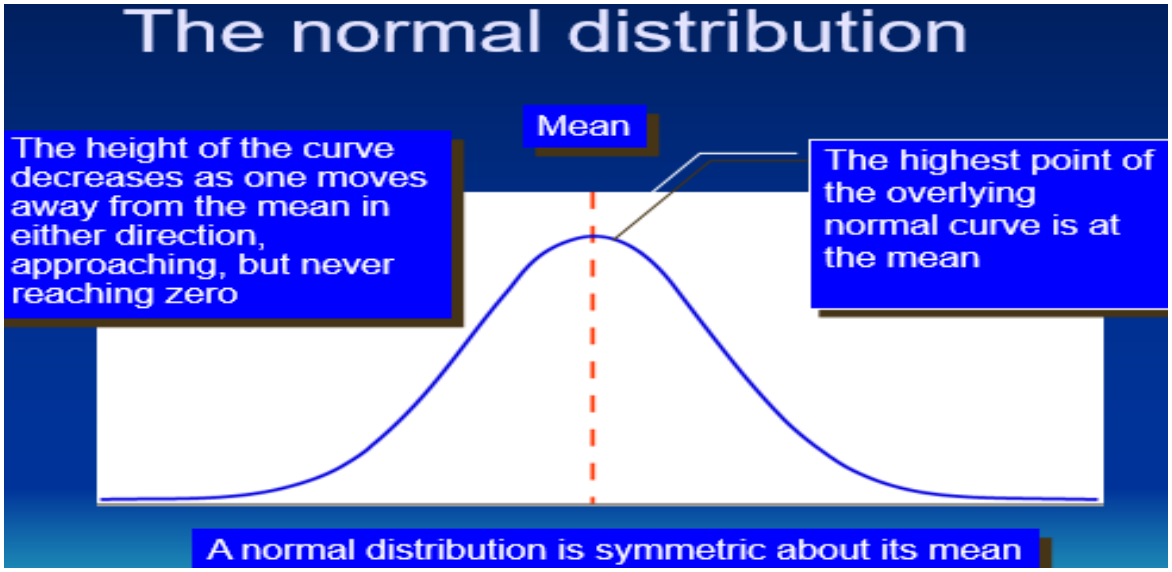



***Figure 6:*** *Frequency polygon (the dotted lines).*

**5- Frequency distribution-curves:**

- If the frequency polygon is smoothed, the frequency distribution curve is obtained. Frequency distribution curve are also obtained by direct plotting of the frequency against the distribution (or simple the measured values) of a certain phenomenon. The shape of the curve differs according to the distribution of the measured phenomenon. Frequency distribution curves are classified into 2 distinct types:


The normal distribution
Mean
The height of the curve decreases as one moves away from the mean in either direction, approaching, but never reaching zero
The highest point of the overlying normal curve is at the mean
A normal distribution is symmetric about its mean

1- *Symmetric frequency distribution curves:* When the curve can be divided into 2 equal parts by a line passing through the center of the distribution (Figure 6).

a) Bell-shaped: and the Majority of measures follow this pattern e.g. heights, weights, etc.…

b) U-Shaped curve: where the data concentrate at both ends of the curve.

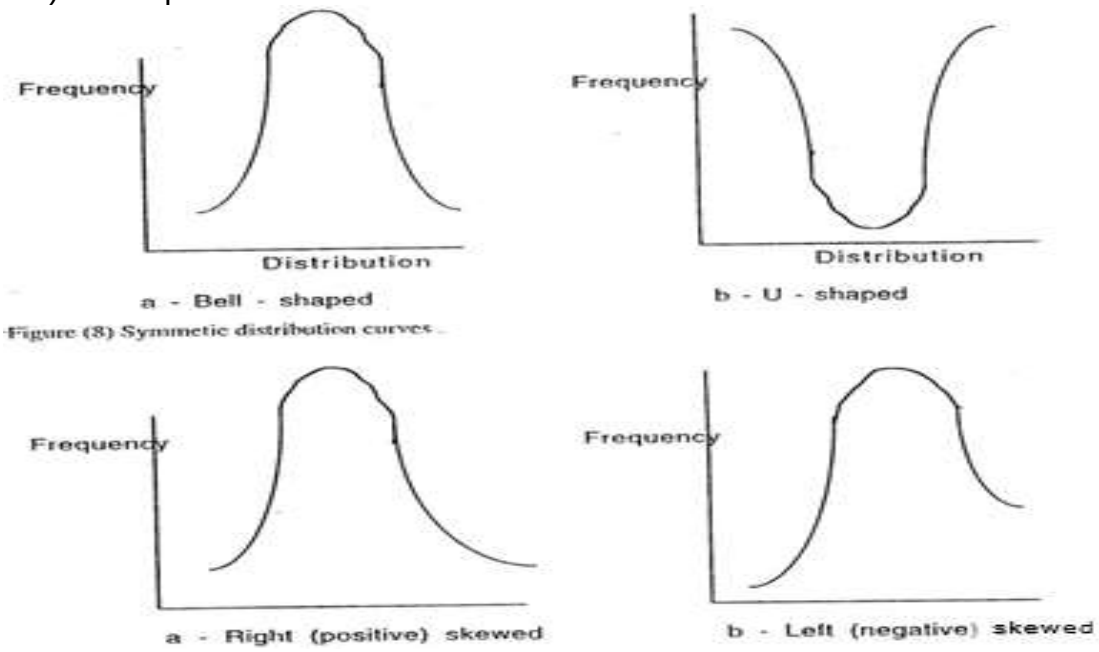



Figure (8) Symmetic distribution curves

***Figure 7:*** *Frequency distribution curves (symmetric and asymmetric).*

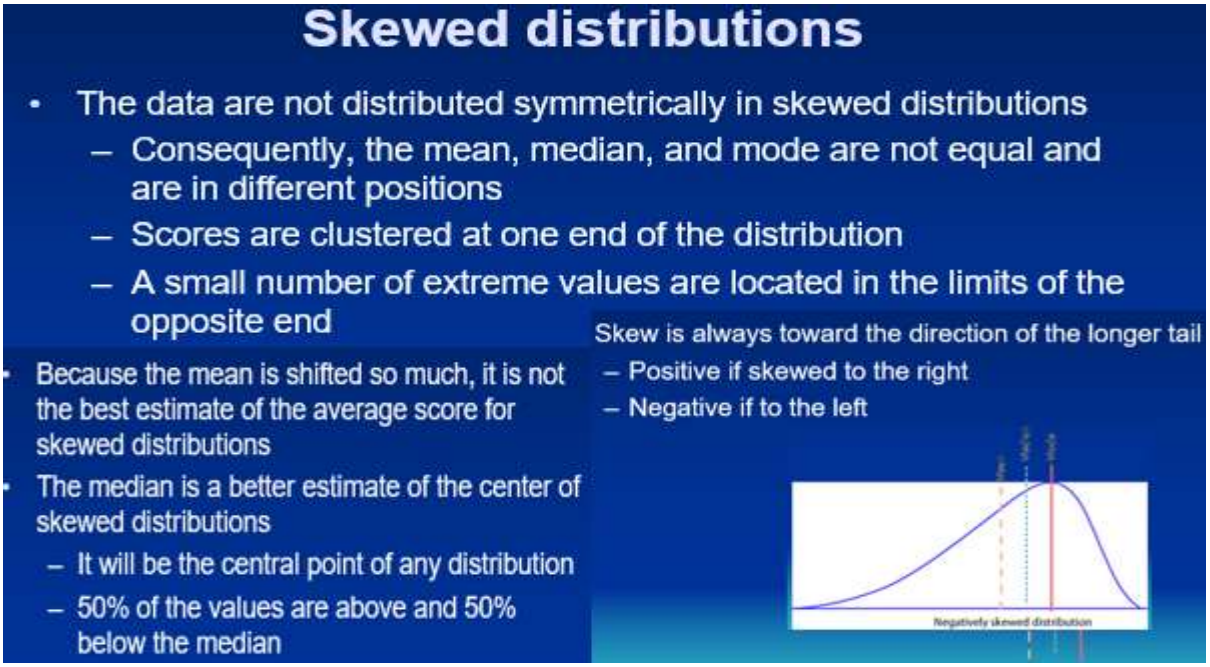
*2- Asymmetric or skewed distribution curves:*

*When the distributions extend further in one direction than the other (Figure 7).*

**Asymmetric distribution curves can be differentiated into:**

a) Positively skewed curves: When the curves skewed to the right and is caused by the extremes in the higher values shifting the curve towards the right.

b) Negatively skewed curves: When the curves skewed to the left, and this is less common type and is caused by extremes in the lower values which distort the curve towards the left.


Skewed distributions
The data are not distributed symmetrically in skewed distributions
– Consequently, the mean, median, and mode are not equal and are in different positions
– Scores are clustered at one end of the distribution
– A small number of extreme values are located in the limits of the opposite end
Because the mean is shifted so much, it is not the best estimate of the average score for skewed distributions
The median is a better estimate of the center of skewed distributions
– It will be the central point of any distribution
– 50% of the values are above and 50% below the median
Skew is always toward the direction of the longer tail
– Positive if skewed to the right
– Negative if to the left
Negatively skewed distribution

**6- Line graph Example:**

The mean blood pressure of 4 groups of cats was measured after an I.V. injection of papaverine in different doses, the following results were obtained

| Dose of papaverine (mg/kg) | 0.5 | 1 | 2 | 3 |
|---|---|---|---|---|
| Mean blood pressure (mm./Hg) | 120 | 80 | 70 | 65 |

-Represent these data graphically.

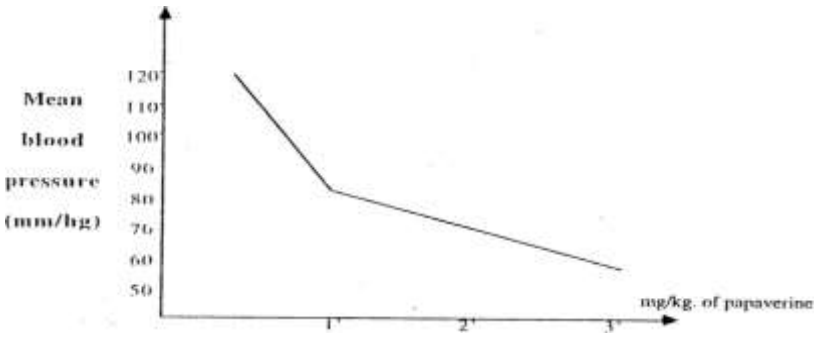



*Figure: Effect of papaverine on the mean blood pressure or cats.*

**Example:**

The sleeping time of mice was determined after an I.V. Injection of different doses of thiopental to 2 groups of mice. The first group was pretreated with chlorpromazine while the second group was control. The following results were obtained.

| Dose of Thiopental (mg/kg) | Pretreated mice | Control mice |
|---|---|---|
| 40 | 15 | 10 |
| 50 | 18 | 12 |
| 60 | 22 | 20 |
| 70 | 25 | 22 |

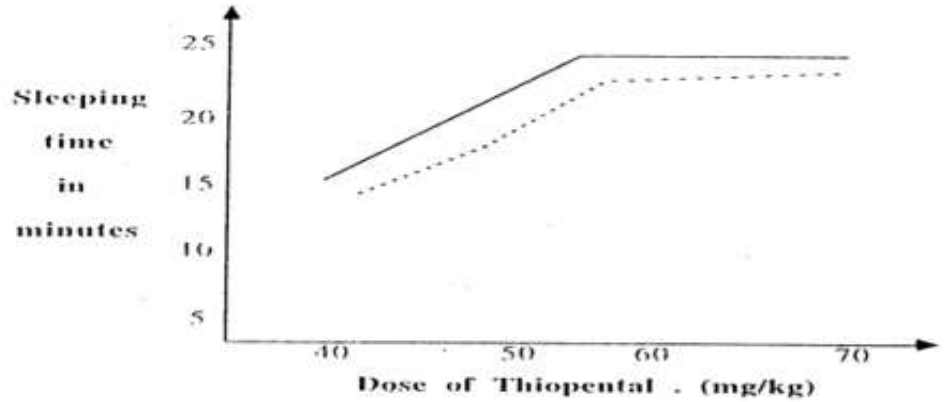



***Figure 8:*** *Effect of pretreatment of mice with chlorpromazine on the sleeping time of mice after I.V. injection of thiopental in different doses.*

_____ Mice pretreated with chlorpromazine.

------- Mice not pretreated with chlorpromazine.

## 7- Scatter or Dot Diagram:

-It is prepared after cross tabulation in which frequencies of at least to variables have been cross classified.

It is a graphic presentation, made to show the nature of correlation between two variable characters X and Y in the same persons or groups such as height and weight in men aged 20 years hence it is also called “Correlation diagram”.

-The characters are read on the base (height) and vertical (weight) axes and the perpendicular drawn from these readings meet to give one scatter point. Varying frequencies of the characters give a number of such points or dots that show a scatter.A line is drawn to show the nature of correlation at a glance as in the following figures.

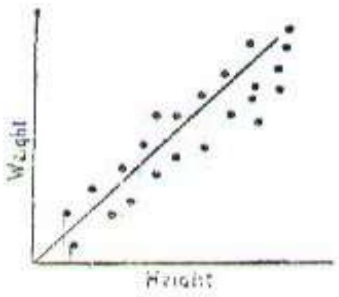



***Figure 10:*** *Scatter diagram showing positive correlation.*

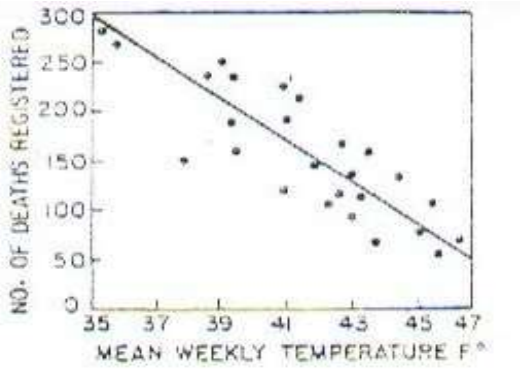



***Figure 11:*** *Scatter diagram showing negative correlation.*

## 9- Pictogram (Picture diagram):

- It is a diagrammatic representation of frequencies, R.F. or magnitude using symbols e.g. drawing a picture.

**Example:** The pictures are drawn in horizontal lines. Each picture indicates a unit of 10, 20, 100, 1000 or 10,000 happenings. Number of pictures in each row gives a quickidea of relative frequency of an attribute. Fraction of a unit has to be ignored, though half (0.5) may be indicated by half picture as in Fig. (3), which gives incidence of choleracases in 4 cities.

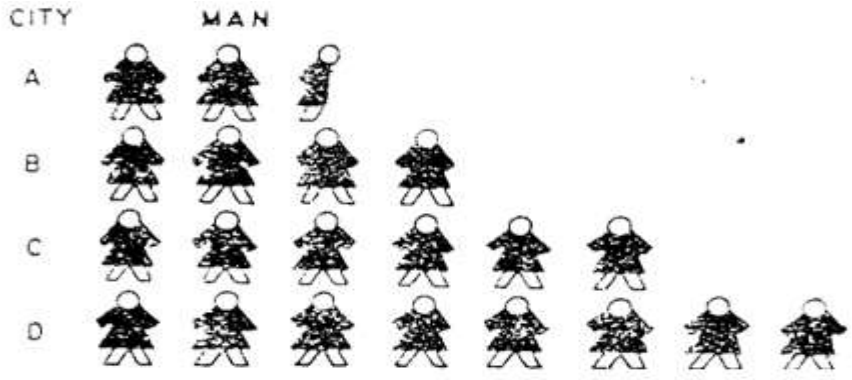



***Figure 12:*** *Pictogram of cholera in the 4 cities, each man's figure indicates 10 cases.*

# Applications of Biostatistics

**Medical statistics:** Go under different names when applied in different fields such as:

1) **Health statistics:** In public health or community health.
2) **Vital statistics:** In demography pertaining to vital events of birth, marriage and death.
3) **In medicine:** It related to the study of defect, injury, disease, efficacy of drugs, serum level and line of treatment etc…

## Application and uses of Biostatistics as a science

1) In physiology and anatomy:

A) To define what is normal or healthy in a population and to define limits of normality in variables as weight and pulse rate. The mean pulse rate is 72/min but up to limits it may be normal on either side of mean has to established with certain appropriate techniques.
B) To find the difference between means and proportions of normal at 2 places or in different periods.
C) To find the correlation between two variables X and Y such as height and weight.

2) In pharmacology:

A) To find the action of a drug. A drug is given to animals or human to sees whether the changes produced are due to the drug or by chance.
B) To compare the action of two different drugs or two successive dosage of the same drug.
C) To find the relative potency of a new drug with respect to a standard drug.

3) In medicine:

A) To compare the efficacy of a particular drug, operation or line of treatment. For this the percent cured, relieved or died in experiment and control groups is compared and find if the difference is due to chance or not.
B) To find an association between two factors such as smoking and lung cancer.
C) To identify signs and symptoms of a disease or syndrome e.g. cough in typhoid.

## Notations for a population and sample values:

** A Statistic:*

It is a descriptive measure computed from the data of a sample.

** A Parameter:*

It is a descriptive measure computed from the data of a population.

Since it is difficult to measure a parameter from the population, a sample is drawn of size n, whose values are $\chi_1$ , $\chi_2$ , …, $\chi_n$. From this data, we measure the statistic.

- Roman letters are used for statistics of sample and Greek for parameters of population, common notations are:

| Value | Sample | Population |
|---|---|---|
| 1- Mean | $\bar{X}$ | $\mu$ |
| 2- Standard deviation | S | $\delta$ |
| 3- Variance | $S^2$ | $\delta^2$ |

**Measures of central tendency (Averages)**

Is a measure which indicates the location of the middle of the data.
The three most commonly used measures of central tendency are:

1- Arithmetic Mean ($\bar{X}$).
2- Median (Middle).
3- Mode.

**1- Mean:** It is the average of the data.

$$\bar{X} = \Sigma X / N$$

$\bar{X}$ = mean.

X = individual result of observation. N = numbers of observation.

Properties of the Mean:

- Uniqueness. For a given set of data there is one and only one mean.
- Simplicity. It is easy to understand and to compute.
- Affected by extreme values. Since all values enter into the computation.

**Example:** The blood glucose level of 6 normal rabbits was found to be 80- 90- 95- 85- 73- 87 calculate the mean blood glucose.

**Solution:**

$\bar{X}$ = (80+90+95+85+73+87) /6 =85

a value that is not representative of the set of data as a whole.

**Example:** Calculate the mean of the following data.

41- 36- 22- 29- 47.

**Solution:**

$\bar{X}$ = (41+ 36+22+ 29+ 47)/ 5 = 35

The algebraic sum for deviations of measurements of a series of observation from their mean =zero.

| X | $\bar{X}$ | D |
|---|---|---|
| 41 | 35 | +6 |
| 36 | 35 | +1 |
| 22 | 35 | -13 |
| 29 | 35 | -6 |
| 47 | 35 | +12 |
| | | Zero |

1- Calculation of $\bar{X}$ from frequency distribution table:

| Systolic B.P. | Midpoint (X) | Frequency (F) | FX |
|---|---|---|---|
| 110- | 112.5 | 3 | 337.5 |
| 115- | 117.5 | 10 | 1175.0 |
| 120- | 122.5 | 22 | 2695.0 |
| 125- | 127.5 | 9 | 1147.5 |
| 130- | 132.5 | 6 | 795.0 |
| | | | 6150 |

$\Sigma.F.X = 6150$

$N = 50$

$X = \Sigma.F.X / N = 6150 / 50 = 123$

**2- Median:**

When ordering the data, it is the observation that divide the set of observations into two equal parts such that half of the data are before it and the other are after it.

* If n is odd, the median will be the middle of observations. It will be the $(n+1)/2^{th}$ ordered observation.

When n = 11, then the median is the $6^{th}$ observation.

* If n is even, there are two middle observations. The median will be the mean of these two middle observations. It will be the mean of the [ $(n/2)^{th}$ , $(n/2 +1)^{th}$ ]ordered observation.

When n = 12, then the median is the $6.5^{th}$ observation, which is an observation halfway between the $6^{th}$ and $7^{th}$ ordered observation.

## Properties of the Median:

- Uniqueness. For a given set of data there is one and only one median.
- Simplicity. It is easy to calculate.
- It is not affected by extreme values as is the mean.

**Example:** The heights of 5 boys is the following.144- 102- 123- 162- 104

**Solution:**

Arrangement: 102- 104- 123- 144- 162

Median position = (N + 1) / 2= 5+1/ 2 = 3

The third value = 123

**Example:** Liver microsomal protein concentration of 8 animals were 23 - 16- 18- 20-16- 19- 21- 22 mg/g liver weight. Determine the median.

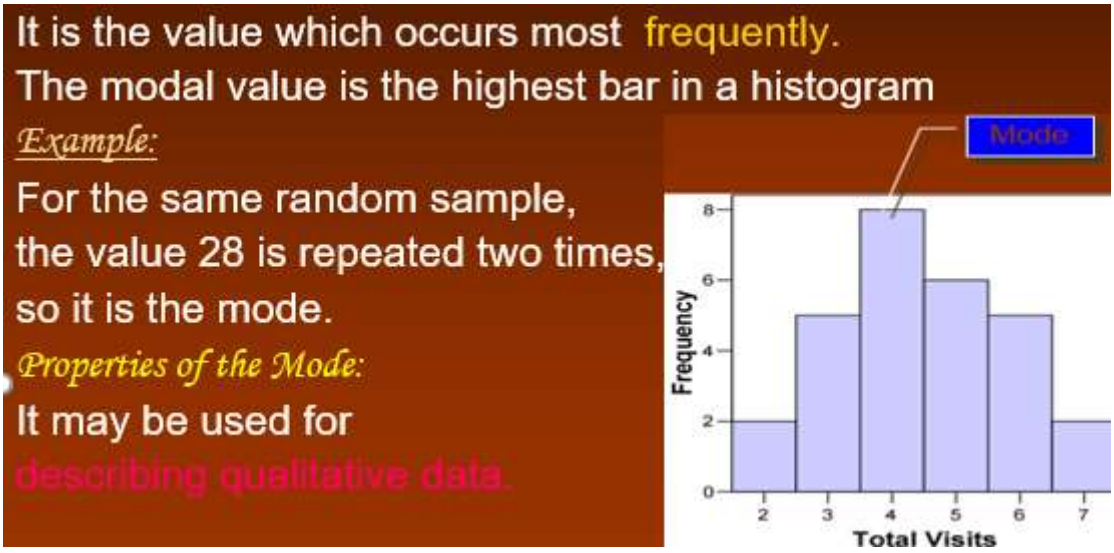
**Solution:**

Arrangement: 16- 16- 18- 19- 20- 21- 22- 23

Median position = 8+1 / 2 = 4.5

Median = mean value of the 4th + $5^{th}$ = 19.5

**3- Mode:**




**Example:** Determine the mode of the following set of data.6- 5- 3- 7- 6- 4- 8

**Mode:** = 6

**Example:** Determine the mode of the following set of data. 6- 5- 3- 7- 3- 6- 8- 4

| Number | | F |
|---|---|---|
| 6 | → | 2 |
| 5 | → | 1 |
| 3 | → | 2 |
| 7 | → | 1 |
| 8 | → | 1 |
| 4 | → | 1 |

The modes are (6 and 3).

**The mode of F.D.T.**

**Example:**

| Heights (cm) | F |
|---|---|
| 120- | 10 |
| 125- | 30 |
| 130- | 40 |
| 135- | 15 |
| 140- | 5 |

**Solution:**

| Heights (cm) | Midpoint | F |
|---|---|---|
| 120- | 122.5 | 10 |
| 125- | 127.5 | 30 |
| 130- | 132.5 | 40 |
| 135- | 137.5 | 15 |
| 140- | 142.5 | 5 |

The mode is the midpoint of the class having the highest F

i.e. the mode is 132.5

## Measures of central tendency

- Each of the three methods of measuring central tendency has certain advantages and disadvantages
- Which method should be used?
  - It depends on the type of data that is being analyzed
  - e.g., categorical, continuous, and the level of measurement that is involved

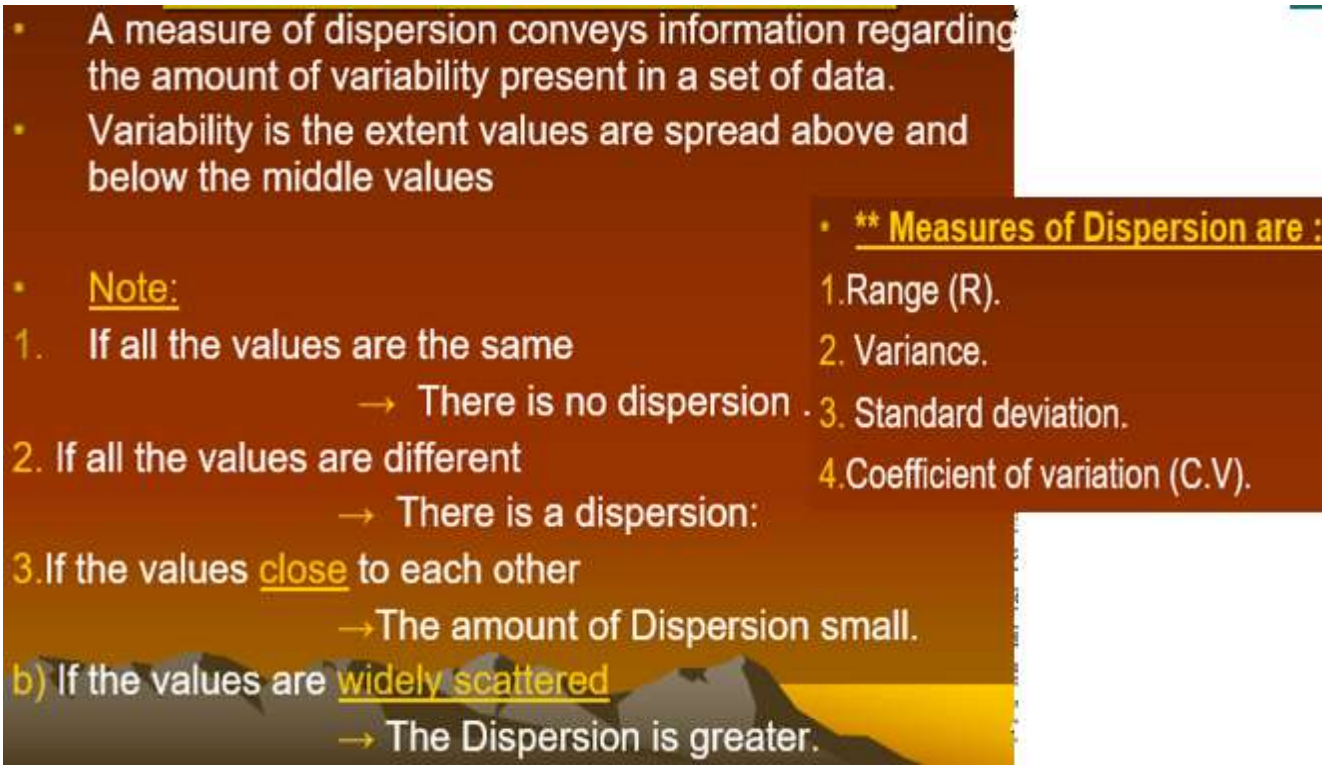
## Levels of measurement

NOIR is a mnemonic to help remember the names and order of the levels of measurement

**N**ominal **O**rdinal Interval **R**atio

| Measurement scale | Permissible mathematic operations | Best measure of central tendency |
|---|---|---|
| Nominal | Counting | Mode |
| Ordinal | Greater or less than operations | Median |
| Interval | Addition and subtraction | Symmetrical – Mean<br>Skewed – Median |
| Ratio | Addition, subtraction, multiplication and division | Symmetrical – Mean<br>Skewed – Median |

# Measures of dispersion


A measure of dispersion conveys information regarding the amount of variability present in a set of data.
Variability is the extent values are spread above and below the middle values
Note:
1. If all the values are the same
→ There is no dispersion
2. If all the values are different
→ There is a dispersion:
3.If the values close to each other
→The amount of Dispersion small.
b) If the values are widely scattered
→ The Dispersion is greater.
** Measures of Dispersion are :
1.Range (R).
2. Variance.
3. Standard deviation.
4.Coefficient of variation (C.V).

Because the arithmetic mean ($\bar{X}$) does not give all information about the data, we look for other measures.

**Example:** We have two sets of data having the same mean ($\bar{X}$), but there is a great deviation between them.

| $X_1$ | $X^-$ | D |
|---|---|---|
| 18 | 19 | 1 |
| 17 | 19 | 2 |
| 22 | 19 | 3 |
| 19 | 19 | - |
| $\bar{X} = \frac{\Sigma d}{4} = 19$<br>$\Sigma d = 1 + 2 + 3 = 6$ | | |

| $X_1$ | $X^-$ | d |
|---|---|---|
| 26 | 19 | 7 |
| 11 | 19 | 8 |
| 17 | 19 | 2 |
| 22 | 19 | 3 |
| $\bar{X} = \frac{\Sigma d}{4} = 19$<br>$\Sigma d = 7 + 8 + 2 + 3 = 20$ | | |

So the deviation of the second set of data is mush greater than in the first one.

**1- Range (R):**

$$R = X_{max} - X_{min}$$

**Example:** Determine the range of the following is sets of data.

1: 18- 17- 22- 19

2: 26- 11- 17- 22

$R_1 = 22 - 17 = 5.$

$R_2 = 26 - 11 = 15.$

**Disadvantages:** It depend only on two values.

**2- Average Deviation (A.D.):**

$$A.D. = \Sigma d / N$$

**N.B.:** Σd = is calculated regardless of the sign in the previous example.

$A.D._1 = 6/4 = 1.5$

$A.D._2 = 20/4 = 5$

This means that deviation in the second set is greater than the first.

**3- Variance (S2):**

It is the mean squared deviations of the values from their mean divided by the number of observation minus 1.

$$S^2 = \Sigma (X-\bar{X})^2 / N-1 = \Sigma d^2/N-1$$

$S^2$ = Variance.

X = Individual value.

$\bar{X}$ = Mean.

N = Number of observation.

d2 = Squared deviation of the data from their arithmetic mean.

**Example:**

| Value X | Mean $\bar{X}$ | Deviation d | Squared $d^2$ | $\Sigma d^2$ |
|---|---|---|---|---|
| 26 | 23 | 3 | 9 | 20 |
| 22 | 23 | 1 | 1 | |
| 24 | 23 | 1 | 1 | |
| 20 | 23 | 3 | 9 | |

$S^2 = = \Sigma (X-\bar{X})^2 / N-1 = \Sigma d^2/N-1$

$S^2 = 20/3 = 6.67$

## 4- Standard Deviation: S or S.D.

$S = \sqrt{\text{Variance}}$

$$S.D. = \sqrt{\frac{\Sigma d^2}{n-1}}$$

- It is the most commonly used measure of variability.
- It is the square root of the variance.

N – 1 = called the degree of freedom (D.F.).

- If the observation is sufficiently large (more than 30) we can use N instead of N- 1 because the product dose not differ significantly in this case.
- In case of much variation the S.D. is high i.e. the higher S.D. the more spread out measured.

.

Valid mean = or > 2.5 S.D.

**Example:** Calculate the standard deviation of the following data 20- 22- 24- 26

| X | $\bar{X}$ | d | $d^2$ | $\Sigma d^2$ |
|---|---|---|---|---|
| 20 | 23 | 3 | 9 | 20 |
| 22 | 23 | 1 | 1 | |
| 24 | 23 | 1 | 1 | |
| 26 | 23 | 3 | 9 | |

$S.D. = \sqrt{\Sigma d^2 / n-1} = \sqrt{20/3} = \sqrt{6.67} = 2.58$

**N.B.:**

Valid mean $\bar{X}$ = or > 2.5 S.D.

2.5 X 2.58 = 6.45

$\bar{X}$ = 23 > 6.45

i.e. the mean is valid.

## Standard Deviation and Variance

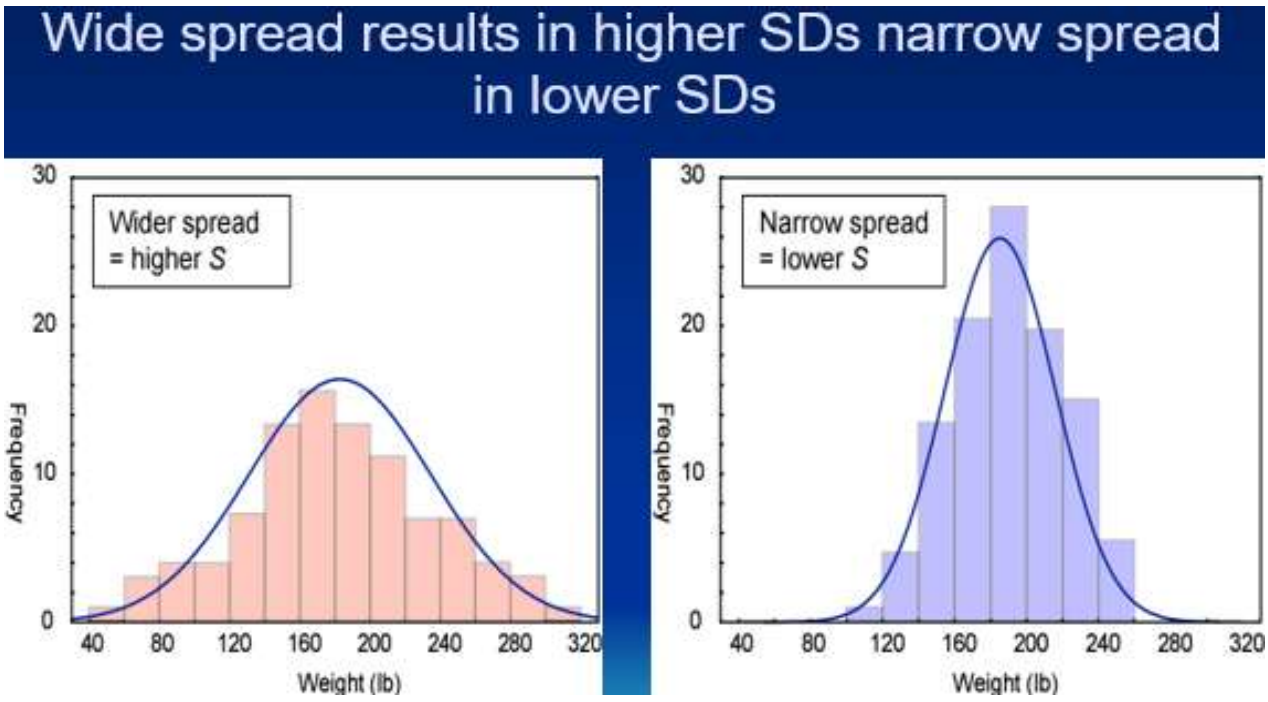
The standard deviation describes the "average distance" of data values from their mean.

| | Xi (individual measurement) | Xi-$\bar{X}$= $d_i$ (indiv – mean measurement) | $(di)^2$ |
|---|---|---|---|
| | 10 | 0.8 | 0.64 |
| | 12 | 1.2 | 1.44 |
| | 8 | -1.2 | 1.44 |
| | 14 | 3.2 | 10.24 |
| | 10 | 0.8 | 0.64 |
| TOTAL | 54 | | 14 |

Mean = 54/ 5 = 10.8

Variance = $S^2$= $(di)^2$/n-1= 14/4 = 3.6

Standard deviation= S= $\sqrt{3.6}$= 1.90




### 5- Coefficient of variation:

- Is a measure use to compare the dispersion in two sets of data which is independent of the unit of the measurement

$$C.V. = \frac{S.D.}{\bar{X}} \times 100$$

$$= \frac{S.}{\bar{X}} \times 100$$

It is used to facilitate the comparison e.g. if we have a set of data whose unit in kg, this cannot be compared directly with another set of data in cm.

**Example:** 2 drugs were bio-assayed for their action on the body weight of 2 groups of rats. The mean body weight of the first group was 50 gm with a S.D. 4.5, andthe second group was 75 gm with S.D. 5.5 compare the degree of dispersion of those figures around their means in the 2 groups.

**Solution:**

C.V. of 1st group = $\frac{4.5}{50}$ X 100 = 9

C.V. of 2nd group = $\frac{5.5}{7.5}$ X 100 = 7.33

This means that the body weight of 1st group is more dispersed around their mean than the second group.

**Example:** In screening of a new drug by the isolated rabbit heart, the mean valueof the heart rate was 70 beat/min and the S.D. was 9.5.

The % of catecholamine uptake was 46.21 with S.D. of 4.24 compare the variabilityof the result of the H.R. and catecholamine uptake.

**Solution:**

C.V. H.R. = 9.5/70 X 100 = 13.7

C.V. E.D = 4.24/46.21 X 100 = 9.17

This indicate that the result of H.R. is more variable or more dispersed around the meanthan the second experiment.

- Suppose two samples of human males yield the following data:
- We wish to know which is more variable.

| | Sampe1 | Sample2 |
|---|---|---|
| Age | 25-year-olds | 11year-olds |
| Mean weight | 145 pound | 80 pound |
| **Standard deviation** | 10 pound | 10 pound |

## 6- Standard Error (S.E)

$$S.E. = \sqrt{\frac{\Sigma d^2}{n\,(n-1)}} = \frac{S.D}{\sqrt{N}} = \sqrt{\frac{\Sigma d^2}{N-1}} \Big/ \sqrt{N}$$

- It is the standard deviation of the mean.
- It describes the variability of the mean.
- Valid mean = or > 10 S.E.

**Example:** Calculate the S.E. of the following data.

20- 22- 24- 26

From previous example S.D. 2.58

$$S.E. = \frac{S.D}{\sqrt{N}} = \frac{2.58}{\sqrt{4}} = 1.29$$

$\bar{X} = 23$

10 S.E. = 10 X 1.29 = 12.9

i.e. the mean is valid

23 > 12.9

## Tests of significance

Tests of significance are used to determine whether the difference between two or more values is significant or not.

**A significant difference:** Means that it will still persist when the experiment is repeated several times and will not be zero at any time.

**A Non-significant difference:** Means that the difference can occur normally and that difference can occur by chance and not due to an external factor such as the effect of a drug or due to a disease. The tests of significance depend essentially on the fact that there are factor or items which are calculated to determine the limits of the normal variations that occur by chance.

**These factors** are given in tables and their values depend on the number of experiments in the series of observations.

**Examples** of these factors or the Z, t, $X^2$ and F values.

These values are calculated from the results of the experiment and the calculated value is compared with the tabulated value.

**If the calculated value** is < the tabulated value the difference is non-significant.

**On the other hand.**

**If the calculated value** is > the tabulated one the difference is significant.

## Levels of significance

When applying the test of significance, we should be sure of our results or decision by a suitable % e.g. we want to be sure that the difference is significant by 99% or by 95% or 99.9%.

This means that we leave (1% or 5% or 0.1%) a chance that the difference is not significant.

When we choose the level of significance to be 99%, this means that the that thedifference is significant by 99% and the probability that the difference is not significant is 1% or 0.01. So, termed as P = 0.01

### Null Hypothesis

It is the hypothesis of no significant difference between the two mean values. We assume that the observed difference is due to chance. We start to analyze the data statistically and calculate suitable factor (t, F, $X^2$ etc…). If as a result of analysis, we found that the difference was not significant we reject the "Null hypothesis".

In bioassay of drug, we choose a dose of the "standard" and determine its effect in the animal and we assume that the "test" has the same concentration like the standard according to the Null hypothesis.

After recording the data and its statistical analysis, we may accept or reject the Null hypothesis depending on the results of the analysis.

**Make decision to accept or reject null hypothesis.**

Slide 21

| Reject the null hypothesis | Accept the null hypothesis |
|---|---|
| Calculated statistic value > the tabulated value at p = 0.05 | Calculated statistic value < the tabulated value at p= 0.05 |
| There is a significant difference | There is No significant difference |
| The difference seen is a real difference and will persist when the experiment is repeated | The difference seen is not real and can occur by chance |

**Confidence limits OR Fiducial limits**

They are the maximal and minimal value which is reached by the mean value at a certain level of significance.

In other words, it is the range within which the mean value can move upward or downward at a certain level of significance.

C.L. = $\bar{X} \pm t.\ S.E.$

C.L. = $\bar{X} + t.\ S.E.$ is upper limit

C.L. = $\bar{X} - t.\ S.E.$ is lower limit

C.L. = Confidence limit.

$\bar{X}$ = Mean.

S.E. = standard error.

t = obtained from the table using P and D.F.

D.F. = Degree of freedom = N-1.

**Example:** The effect of certain drug was studied on the body weight of 10 mice. The following results were obtained

17- 23- 20- 20- 20- 23- 20- 35- 20- 23

Find out the confidence limits of the mean of this group of mice at a level of significance 95% (P = 0.05).

**Solution:** C.L. = $\bar{X} \pm t.\ S.E.$

$\bar{X} = \Sigma X/N = 22.1$

| X | $\bar{X}$ | D | $d^2$ | $\Sigma d^2$ |
|---|---|---|---|---|
| 17 | 22.1 | 5.1 | 26.01 | 216.90 |
| 23 | 22.1 | 0.9 | 0.81 | |
| 20 | 22.1 | 2.1 | 4.41 | |
| 20 | 22.1 | 2.1 | 4.41 | |
| 20 | 22.1 | 2.1 | 4.41 | |
| 23 | 22.1 | 0.9 | 0.81 | |
| 20 | 22.1 | 2.1 | 4.41 | |
| 35 | 22.1 | 12.9 | 166.41 | |
| 20 | 22.1 | 2.1 | 4.41 | |
| 23 | 22.1 | 0.9 | 0.81 | |

S.E. = $\sqrt{\ }\Sigma d^2/N\ (N-1) = \sqrt{\ }216.9/(10X9) = \sqrt{\ }216.9/90 = \sqrt{\ }2.41 = 1.55$

Degree of freedom = N-1 = 10-1 = 9

From the table we find "t" under d.f = 9 at P = 0.05

From the table t = 2.26

C.L. upper = $\bar{X}$ + S.E. t = 22.1 + 1.55 X 2.26 = 22.1 + 3.5 = 25.6

C.L. lower = 22.1 – 3.5 = 18.6

- This mean that 95% confidence limit of the mean are ranging from 18.6 to 25.6.
- Also the confidence limits of the mean expressed as follow 22.1 ± 3.5 at P = 0.05

**Comment:**

(C.L.) confidence limits of the mean inform that the mean of a given experiment is not a fixed value, but it is a range that has lower and upper values or limits.

- The 95% C.L. means that if the experiment at hand is theoretically repeated 100 times, it is possible to obtain ($\bar{X}$) with its calculated lower and upper (C.L.) in 95% of the experiment. Only 5 experiments will give results deviating from those obtained in the practically conducted experiment.
- The 99% (C.L.) has the same role except that only 1 one experiment will give rise to deviate from those of the practically carried out experiment.

**Example:** In a biological assay of the potency of a certain hormone injection, the following values were obtained for individual assays.

| |
|---|
| 0.738 |
| 0.760 |
| 0.803 |
| 0.817 |
| 0.870 |
| 0.889 |

Find out the C.L. in this assay at P = 0.05 t = 2.571

**Solution:**

$\bar{X}$ = 0.813

S.E. = 0.024

C.L. = $\bar{X}$ ± S.E. t.

C.L. = 0.813 + 0.024 X 2.57 = 0.8755

0.813 - 0.024 X 2.57 = 0.7521

## Tests of significance

### (1) Normal deviate test

It is a procedure through which the truth of the Null hypothesis (hypothesis of difference) is investigated by examining certain value of the test.

The decision is taken to reject the Null hypothesis i.e. there is a significant difference at certain level of significance.

This test is used to determine whether the ($\bar{X}$) mean of a small set of result differ significantly from the universal mean ($\mu$).

There are two conditions:

A) If the S.D. of the population is known.

B) If the S.D. of the population is Unknown.

### A) If the S.D. of the population is known.

$$Z = \bar{X} - \mu / (S.D./\sqrt{n})$$

Z = Normal deviate.

$\bar{X}$ = Mean of the sample.

$\mu$ = Mean of the population.

n = number of samples.

S.D. = standard deviation of the population.

The calculated "Z" value is compared with the tabulated value at P<0.05 and D.F. at $\alpha$. "Z" is usually = 1.96

**Example:** The purity of a product was checked by analysis of a sample from each batch. The analysis of nitrogen averaged over a long period was 4.02% with S.D. 0.06%. Four successive batches in one week has a mean nitrogen content of4.2%. Does this indicate any significant change in the composition of this product?

**Solution:**

$$Z = \bar{X} - \mu / (S.D./\sqrt{n}) = 4.20 - 4.02/(0.06/\sqrt{4}) = 0.18/(0.06/2) = 6$$

Calculated "Z" = 6

Tabulated "Z" = 1.96 at $\alpha$.

Z cal > Z tab

i.e. the Null hypothesis is rejected and there is significant difference.

**OR** The $\bar{X}$ (Mean of the sample) is not representative of the $\mu$ (Mean of the population).

**B) If the S.D. of the population is Unknown.**

The calculated "Z" is compared with the tabulated are from t-distribution table at D.F. = (n-1) and P <0.05

The Null hypothesis is accepted if Z value is less than the tabulated one. i.e. no significant difference.

**Example:** The mean clotting time of 11 individual is 11.6. If this sample is taken from a population with a mean of 10. the estimated S.D. of the sample is 1.99 See if there is a real difference between the two means.

**Solution:**

Z calculated= $11.6-10 / (1.99/\sqrt{11}) = 1.6/ (1.99/\sqrt{11}) = 1.6/ (1.99/3.3) = 2.7$

Tabulated "Z" at D.F. = n-1 = 11-1 = 10

Equal **2.23**

So the calculated > tabulated i.e. there is a significant difference between the two means.

**Example:** The mean active content of 10 tablets from one batch was 139 mg/tab with S.D. 40. Is this value significant from the expected content of the tablet (130 mg)?

| D.F. | Z |
|---|---|
| 9 | 2.26 |
| 10 | 2.22 |
| 20 | 2.09 |
| $\alpha$ | 1.96 |

**Solution:**

Z = 0.72

"Z" at D.F. (9) = 2.26

Calculated < tabulated i.e. no significant difference the Null hypothesis is accepted.

## 2- Student's t-test

### A) Independent or unpaired

It is the most commonly used test in graded pharmacological assay. It is used to determine the significance of the difference between two mean values. Each mean value has its S.E. and S.D.

The test depend on calculation of the "t" value depending on the two mean values and their S.E., as well as, the number of experiments in each series of observations.

**t = difference/ S.E. of the difference=** $X_1^- - X_2^- / (\sqrt{S.E._1^2 + S.E._2^2})$

**Example:** The bioassay of epinephrine (EN) was carried out on the dog blood pressure using matching technique. The heights of contraction given by the submaximal doses of the standard EN were [8- 6- 6.5- 7- 7.5] cm.

The heights of contraction given by the same dose of the test preparation were [6- 5- 5.5- 6- 5] cm.

Find out if there is a real difference between the standard and the test.

| Standard | | | |
|---|---|---|---|
| X | $X^-$ | d | $d^2$ |
| 8 | 7 | 1 | 1 |
| 6 | 7 | 1 | 1 |
| 6.5 | 7 | 0.5 | 0.25 |
| 7 | 7 | 0 | 0 |
| 7.5 | 7 | 0.5 | 0.25 |
| | | | 0.50 |

| Test | | | |
|---|---|---|---|
| X | $X^-$ | d | $d^2$ |
| 6 | 5.5 | 0.5 | 0.25 |
| 5 | 5.5 | 0.5 | 0.25 |
| 5.5 | 5.5 | 0 | 0 |
| 6 | 5.5 | 0.5 | 0.25 |
| 5 | 5.5 | 0.5 | 0.25 |
| | | | 1.00 |

$S.E_1 = \sqrt{\Sigma d_2 / n(n-1)} = 0.35$ $S.E_2 = 0.22$

$S.E_1^2 = 0.123$ $S.E_2^2 = 0.05$

t calculated= $X_1^- - X_2^- / (\sqrt{S.E._1^2 + S.E._2^2}) = 7-5.5/ \sqrt{0.123 + 0.050} =$

$= 1.5/ \sqrt{0.173} = 3.57$

D.F. = n1 + n2 – 2 = 5 + 5 – 2 = 8

t at D.F.8 = 2.306

T cal > T tab

i.e. there is a significance difference.

**Example:** The effect of amphetamine and chlorpromazine (CPZ) on the body weight of adult rats was measured by injection of each drug in a group of 8 rats. The results were as follows:

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| CPZ | 50 | 45 | 40 | 44 | 35 | 36 | 33 | 37 |
| Amphetamine | 35 | 30 | 25 | 34 | 32 | 32 | 30 | 22 |

Test the significance of the difference at P = 0.05

| D.F. | T |
|---|---|
| 7 | 2.37 |
| 8 | 2.31 |
| 14 | 2.15 |
| 16 | 2.12 |

**Solution:** $t = 40 - 30/(\sqrt{(2.076)^2 + (1.574)^2}) = 10 / \sqrt{4.309 + 2.478}$

$= 10 / \sqrt{6.787} = 10/2.605 = 3.83$

t cal = 3.83

t tab = 2.15

i.e. there is a significant difference.

## B) Paired t-test

$$t = \bar{X_1} - \bar{X_2} / \sqrt{\Sigma (d - \bar{d})^2 / n (n-1)}$$

$\bar{X_1}$ = Mean of the initial values.

$\bar{X_2}$ = Mean of the treated values.

d = Deviation.

$\bar{d}$ = Mean Deviation = $\Sigma d/n$

n = Number of observation.

**Example:** The effect of newly synthetized compound was investigated on the blood sugar level of 5 normal rats. Before treatment the normal blood sugar was found to be [80- 90- 110- 120- 130] mg%.

After injection of the compound, the blood sugar level was found to be [40- 60- 50- 20- 60] mg%.

Find out whether the decrease in the blood sugar is due to chance or due to the actual effect of the drug.

**Solution:**

| Animal # | $X_1$ | $X_2$ | d | $\bar{d}$ | $(d-\bar{d})$ | $(d-\bar{d})^2$ |
|---|---|---|---|---|---|---|
| 1 | 80 | 40 | 40 | 50 | -10 | 100 |
| 2 | 90 | 60 | 30 | 50 | -20 | 400 |
| 3 | 110 | 50 | 60 | 50 | +10 | 100 |
| 4 | 120 | 70 | 50 | 50 | 0 | 0 |
| 5 | 130 | 60 | 70 | 50 | 20 | 400 |
| | $\bar{X}_1 = 106$ | $\bar{X}_2 = 56$ | 250 | | | 1000 |

$\bar{d} = \Sigma d / n = 250/5 = 50$

$\bar{X}_1 = 106$

$\bar{X}_2 = 56$

$t = \mathbf{t = \bar{X}_1 - \bar{X}_2 / \sqrt{\Sigma (d - \bar{d})^2 / n (n-1) = 106-56 / \sqrt{1000 / 5X4}}}$

$= 50/ \sqrt{1000/20} = 50/ \sqrt{50} = 7.07$

D.F. = n–1 = 5–1 = 4

t tab at (D.F. 4) = 2.77

t cal > t tab

i.e. there is a significant difference.

**Example:**

The effect of new compound on the body weight of rats was investigated in 5 rats. The initial weight was [110- 105- 100- 105- 106] gm. and after treatment was [110- 102- 102- 103- 105] gm.

Does the new compound possess a significant reduction on body weight at P < 0.05 ?

| $X_1$ | $X_2$ | d | $\bar{d}$ | $(d-\bar{d})$ | $(d-\bar{d})^2$ |
|---|---|---|---|---|---|
| 110 | 110 | 0 | 0.8 | -0.8 | 0.64 |
| 105 | 102 | 3 | 0.8 | 2.2 | 4.84 |
| 100 | 102 | -2 | 0.8 | -2.8 | 7.84 |
| 105 | 103 | 2 | 0.8 | 1.2 | 1.44 |
| 106 | 105 | 1 | 0.8 | 0.2 | 0.04 |
| $\bar{X}_1$ = 105.2 | $\bar{X}_2$ = 104.4 | $\Sigma d = 4$ | | | 14.80 |

$\bar{d} = \Sigma d/n = 4/5 = 0.8$

$\bar{X_1} = 105.2$

$\bar{X_2} = 104.4$

t calculated= $\bar{X_1} - \bar{X_2} / (\sqrt{\Sigma(d-\bar{d})^2 / n(n-1)}) = 105.2 - 104.2 / \sqrt{14.8/20}$

**= 0.93**

t tab at D.F. 4 = 2.77

t cal < t tab

i.e. there is no significant difference.

**Example:** The effect of new compound on the body weight of rats was investigated in 6 rats. The initial weight was [100- 115- 130- 120- 130- 125] gm. and after treatment was [105- 85- 80- 75- 85- 80] gm.

Does the new compound possess a significant reduction on body weight at P < 0.05 ?

**Solution:**

| $X_1$ | $X_2$ | D | $\bar{D}$ | $(d-\bar{d})$ | $(d-\bar{d})^2$ |
|---|---|---|---|---|---|
| 100 | 105 | -5 | 35 | -40 | 1600 |
| 115 | 85 | 30 | 35 | -5 | 25 |
| 130 | 80 | 50 | 35 | 15 | 225 |
| 120 | 75 | 45 | 35 | 10 | 100 |
| 130 | 85 | 45 | 35 | 10 | 100 |
| 125 | 80 | 45 | 35 | 10 | 100 |
| $\bar{X_1}$ = 120 | $\bar{X_2}$ = 85 | | | | 2150 |

$\bar{d} = \Sigma d/n = +205-5/6 = 210/6 = 35$

$\bar{X_1} = 120$

$\bar{X_2} = 85$

t = $120 - 85 / \sqrt{2150/30} = 35/\sqrt{71.66} = 4.13$

t tab = 2.57 i.e. there is a significant difference

## 3- Chi-Square Test "$X^2$"

This test is very useful in research and is most commonly used when data are in frequencies such as in the number of responses in two or more categories. It can be used with any data which can be reduced to proportions or percentages.

The test involves the calculation of a quantity called chi-square from the Greek letter "chi" X2 and pronounced as "kye".

Usually table is set which joint occurrence of two sets of events e.g. the treatment and outcome of the disease. This called the contingency table [latin, con = together, tanger = to touch] see table ( ).

X2 is also applied to find association between two events occurring together when there are more than two classes or more than two samples table ( 6 ).

Table (6)

| Groups | Outcome of result | | |
|---|---|---|---|
| | Died | Survived | Total |
| A (control) | 10 | 25 | 35 |
| B (drug) | 5 | 60 | 65 |
| | 15 | 85 | 100 |

Here, there are two columns and two raws.

D.F. = (r – 1) (c– 1)

r = raw

c = columns

= (2 – 1) (2 – 1) = 1X1 = 1

**Table (7)**

| Class | Microfiloria positivity | | | |
|---|---|---|---|---|
| | +ve | –ve | Total | %+ve |
| I | 4 | 76 | 80 | 5% |
| II | 20 | 180 | 200 | 10% |
| III | 60 | 440 | 500 | 12% |
| IV | 144 | 576 | 720 | 20% |
| | 228 | 1272 | 1500 | 15.2% |

**D.F. = (r – 1) (c– 1)**

**= (4 – 1) (2 – 1) = 3X1 = 3**

$$X^2 = \Sigma \frac{(O-E)^2}{E}$$

O = Observed frequency.
E = Expected frequency.
$X^2$ = Chi-Square.

Example #1: The effect of a drug in the treatment of certain disease was determined by its administration to 70 patients. 10 patients only died. If, it is known from past experience that the percent mortality of that disease was 30%. Test whether the drug has a real curative effect on the disease at P < 0.05?

Solution: Expected mortality = = 21 patients.

| | Died | Survive | Total |
|---|---|---|---|
| Disease | 21 | 49 | 70 |
| Drug | 10 | 60 | 70 |

D.F. = (r – 1) (c– 1)

= (2 – 1) (2 – 1) = 1X1 = 1

| | O | E | O - E | (O – E)2 | (O – E)2/E |
|---|---|---|---|---|---|
| Survive | 60 | 49 | 11 | 121 | 121/49 = 2.46 |
| Died | 10 | 21 | 11 | 121 | 121/49 = 5.76 |
| | | | | | Total (chi)2 8.22 |

X2 cal = 8.22

X2 tab at D.F. 1 = 3.84

i.e. cal > tab

So there is significant difference and the drug has a real curative effect.

Example **#2:** The effect of solbutamol and a new bronchodilator drug was compared by administration to 60 patients suffering from bronchospasm. Solbutamol was given to 20 patients, only 15 patients were cured. The new drug was given to the rest of the patients, only 18 patients were cured. Doses the effect of the new drug differ significantly from solbutamol at P < 0.05?

**Solution:**

$$\text{Solbutamol} = \frac{15X100}{20} = 75\%$$

$$\text{New drug} = \frac{18X100}{40} = 45\%$$

| Treatment | Cured | Non cured | Total |
|---|---|---|---|
| Solbutamol | 15 | 5 | 20 |
| New drug | 18 | 22 | 40 |
| Total | 33 | 27 | 60 |

D.F. = (r – 1) (c– 1) = (2 – 1) (2 – 1) = 1X1 = 1

$$\text{Expected F} = \frac{\text{Raw T X Column T}}{\text{Grand T}}$$

$$\text{Expected Solbutamol cured} = \frac{20 \text{ X } 33}{60} = 11$$

$$\text{Expected Solbutamol non cured} = \frac{20 \text{ X } 27}{60} = 9$$

Expected New drug cured = $\frac{40 \text{ X } 33}{60} = 22$

Expected New drug non cured = $\frac{40 \text{ X } 27}{60} = 18$

| | O | E | O - E | $(O - E)^2$ | $(O - E)^2/E$ |
|---|---|---|---|---|---|
| Solbutamol cured | 15 | 11 | 4 | 16 | 16/11 |
| Solbutamol non cured | 5 | 9 | 4 | 16 | 16/19 |
| New drug cured | 18 | 22 | 4 | 16 | 16/22 |
| New drug non cured | 22 | 18 | 4 | 16 | 16/18 |

Total $(Chi)^2 = 4.83$

T $chi^2 = 3.84$

There is a significant difference between the two drugs.

## Table of $X^2$

**Probability (P)**

| D.F. | .50 | .10 | .05 | .02 | .01 | .001 |
|---|---|---|---|---|---|---|
| 1 | 0.46 | 2.71 | 3.84 | 5.41 | 6.64 | 10.83 |
| 2 | 1.39 | 4.61 | 5.99 | 7.82 | 9.21 | 13.82 |
| 3 | 2.37 | 6.25 | 7.82 | 9.84 | 11.34 | 16.27 |
| 4 | 3.35 | 7.78 | 9.49 | 11.67 | 13.28 | 18.47 |
| 5 | 4.35 | 9.24 | 11.07 | 13.39 | 15.09 | 20.52 |
| 6 | 5.35 | 10.65 | 12.59 | 15.03 | 16.81 | 22.46 |
| 7 | 6.35 | 12.02 | 14.07 | 16.62 | 18.48 | 24.32 |
| 8 | 7.34 | 13.36 | 15.51 | 18.17 | 20.09 | 16.13 |
| 9 | 8.34 | 14.68 | 16.92 | 19.68 | 21.67 | 27.88 |
| 10 | 9.34 | 15.89 | 18.31 | 21.16 | 23.21 | 29.59 |
| 11 | 10.34 | 17.28 | 19.68 | 22.62 | 24.73 | 31.26 |
| 12 | 11.34 | 18.55 | 21.03 | 24.05 | 26.22 | 32.91 |
| 13 | 12.34 | 19.81 | 22.36 | 25.47 | 27.69 | 34.53 |
| 14 | 13.34 | 21.06 | 23.69 | 26.87 | 29.14 | 36.12 |
| 15 | 14.34 | 22.31 | 24.99 | 28.26 | 30.58 | 37.70 |
| 16 | 15.34 | 23.54 | 26.30 | 29.63 | 32.00 | 39.25 |
| 17 | 16.34 | 24.77 | 27.59 | 30.99 | 33.41 | 40.79 |
| 18 | 17.34 | 25.99 | 28.87 | 32.35 | 34.81 | 42.31 |
| 19 | 18.34 | 27.20 | 30.14 | 33.69 | 36.19 | 43.82 |
| 20 | 19.34 | 28.41 | 31.41 | 35.02 | 37.57 | 45.32 |
| 21 | 20.34 | 29.62 | 32.67 | 36.34 | 38.93 | 46.80 |
| 22 | 21.34 | 30.81 | 33.92 | 37.66 | 40.29 | 48.27 |
| 23 | 22.34 | 32.01 | 35.17 | 38.97 | 41.64 | 49.73 |
| 24 | 23.34 | 33.20 | 36.42 | 40.27 | 42.98 | 51.18 |
| 25 | 24.34 | 34.38 | 37.65 | 41.57 | 44.31 | 52.62 |
| 26 | 25.34 | 35.56 | 38.89 | 42.86 | 45.64 | 54.05 |
| 27 | 26.34 | 36.74 | 40.11 | 44.14 | 46.96 | 55.48 |
| 28 | 27.34 | 37.92 | 41.34 | 45.42 | 48.28 | 56.89 |
| 29 | 28.34 | 39.09 | 42.56 | 46.69 | 49.59 | 58.30 |
| 30 | 29.34 | 40.26 | 43.77 | 47.96 | 50.89 | 59.70 |


*Course Notes by Dr. Hamdi Ghonim - Dr. Hoda Ezzat Kafl* 49

## 4- Analysis of variance (ANOVA)

This test is used to compare more than two means. The calculations depend on the following

1- Calculation of "between groups" = (group means – grand mean)$^2$, for each group and multiplied by the numbers of observations in each group. Then summing for all the group.

$$N_1 (\bar{X_1} - \bar{X}=)^2 + N_2 (\bar{X_2} - \bar{X}=)^2 + N_3 (\bar{X_3} - \bar{X}=)^2 \ldots\ldots\ldots$$

2- Calculation "within groups"

$$= \text{sum of (observation)}^2 - \frac{(\text{sum of observation in group})^2}{\text{N of observation}} = \Sigma X^2 \qquad \frac{\Sigma X)^2}{N}$$

**Example:** The following is the effect of 3 different drugs on certain enzyme activity in rats.

| Drug A | Drug B | Drug C |
|---|---|---|
| 10 | 11 | 9 |
| 14 | 8 | 11 |
| 13 | 12 | 10 |
| 11 | 10 | 10 |
| 12 | | 13 |
| | | 12 |

Do the three drugs differ significantly from each other at P <0.05?

**Solution:**

| $X_1$ | $X_1^2$ | $X_2$ | $X_2^2$ | $X_3$ | $X_3^2$ |
|---|---|---|---|---|---|
| 10 | 100 | 11 | 121 | 9 | 81 |
| 14 | 196 | 8 | 64 | 11 | 121 |
| 13 | 169 | 12 | 144 | 10 | 100 |
| 11 | 121 | 10 | 100 | 10 | 100 |
| 12 | 144 | | | 13 | 169 |
| | | | | 12 | 144 |
| $\Sigma X =$ | 60 | | 41 | | 65 |
| $\bar{X} =$ | 12 | | 10.45 | | 10.83 |
| $\Sigma X^2 =$ | 730 | | 429 | | 715 |

$$\bar{X}= = \frac{60 + 41 + 65}{15} = \frac{166}{15} = 11.07$$

**1- Between groups:**

$N_1 (\bar{X}_1 - X=)^2 + N_2 (\bar{X}_2 - X=)^2 + N_3 (\bar{X}_3 - X=)^2$

$= 5 (12 - 11.07)^2 + 4 (10.25 - 11.07)^2 + 6 (10.83 - 11.07)^2$

$= 5 (0.93)^2 + 4 (0.82)^2 + 6 (0.24)^2$

$= 5 \text{ X } 0.8649 + 4 \text{ X } 0.6724 + 6 \text{ X } 0.0576$

$= 2.6894 + 4.3245 + 0.3456 = 7.359$

D.F. $= 3 - 1 = 2$

Mean $= \frac{7.359}{2} = 3.67$

**2- Within groups:**

429 – = 429 – = 429 – 420.3 = 8.7

730 – = 730 – = 730 – 720 = 10

715 – = 715 – = 715 – 404.2 = 10.8

Total = 8.7 + 10 + 10.8 = 29.5

D.F. = 15 – 3 = 12

Mean square deviation = = 2.46

F = 1.49

F tabulated at D.F. = 2 & 12 = 3.89

i.e. there is a no significant difference between the 3 groups: (non significant).


*Course Notes by Dr. Hamdi Ghonim - Dr. Hoda Ezzat Kafl* 51

## Table of F value

| Determinator degrees of freedom | Numerator degrees of freedom | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
| 1 | 161.4 | 199.5 | 215.7 | 224.6 | 230.2 | 234.0 | 236.8 | 238.9 | 240.5 |
| 2 | 18.51 | 19.00 | 19.16 | 19.25 | 19.30 | 19.33 | 19.35 | 19.37 | 19.38 |
| 3 | 10.13 | 9.55 | 9.28 | 9.12 | 9.01 | 8.94 | 8.89 | 8.85 | 8.81 |
| 4 | 7.71 | 6.94 | 6.59 | 6.39 | 6.26 | 6.16 | 6.09 | 6.04 | 6.00 |
| 5 | 6.61 | 5.79 | 5.41 | 5.19 | 5.05 | 4.95 | 4.88 | 4.82 | 4.77 |
| 6 | 5.99 | 5.14 | 4.76 | 4.53 | 4.39 | 4.28 | 4.21 | 4.15 | 4.10 |
| 7 | 5.59 | 4.74 | 4.35 | 4.12 | 3.97 | 3.87 | 3.79 | 3.73 | 3.68 |
| 8 | 5.32 | 4.46 | 4.07 | 3.84 | 3.69 | 3.58 | 3.50 | 3.44 | 3.39 |
| 9 | 5.12 | 4.26 | 3.86 | 3.63 | 3.48 | 3.37 | 3.29 | 3.23 | 3.18 |
| 10 | 4.96 | 4.10 | 3.71 | 3.48 | 3.33 | 3.22 | 3.14 | 3.07 | 3.02 |
| 11 | 4.84 | 3.98 | 3.59 | 3.36 | 3.20 | 3.09 | 3.01 | 2.95 | 2.90 |
| 12 | 4.75 | 3.89 | 3.49 | 3.26 | 3.11 | 3.00 | 2.91 | 2.85 | 2.80 |
| 13 | 4.67 | 3.81 | 3.41 | 3.18 | 3.03 | 2.92 | 2.83 | 2.77 | 2.71 |
| 14 | 4.60 | 3.74 | 3.34 | 3.11 | 2.97 | 2.85 | 2.76 | 2.70 | 2.65 |
| 15 | 4.54 | 3.68 | 3.29 | 3.06 | 2.90 | 2.79 | 2.71 | 2.64 | 2.59 |
| 16 | 4.49 | 3.63 | 3.24 | 3.01 | 2.85 | 2.74 | 2.66 | 2.59 | 2.54 |
| 17 | 4.45 | 3.59 | 3.20 | 2.96 | 2.81 | 2.70 | 2.61 | 2.55 | 2.49 |
| 18 | 4.41 | 3.55 | 3.16 | 2.93 | 2.77 | 2.66 | 2.58 | 2.51 | 2.46 |
| 19 | 4.38 | 3.52 | 3.13 | 2.90 | 2.74 | 2.63 | 2.54 | 2.48 | 2.42 |
| 20 | 4.35 | 3.49 | 3.10 | 2.87 | 2.71 | 2.60 | 2.51 | 2.45 | 2.39 |
| 21 | 4.32 | 3.47 | 3.07 | 2.84 | 2.68 | 2.57 | 2.49 | 2.42 | 2.37 |
| 22 | 4.30 | 3.44 | 3.05 | 2.82 | 2.66 | 2.55 | 2.46 | 2.40 | 2.34 |
| 23 | 4.28 | 3.42 | 3.03 | 2.80 | 2.64 | 2.53 | 2.44 | 2.37 | 2.32 |
| 24 | 4.26 | 3.40 | 3.01 | 2.78 | 2.62 | 2.51 | 2.42 | 2.36 | 2.30 |
| 25 | 4.24 | 3.39 | 2.99 | 2.76 | 2.60 | 2.49 | 2.40 | 2.34 | 2.28 |
| 26 | 4.23 | 3.37 | 2.98 | 2.74 | 2.59 | 2.47 | 2.39 | 2.32 | 2.27 |
| 27 | 4.21 | 3.35 | 2.96 | 2.73 | 2.57 | 2.46 | 2.37 | 2.31 | 2.25 |
| 28 | 4.20 | 3.34 | 2.95 | 2.71 | 2.56 | 2.45 | 2.36 | 2.29 | 2.24 |
| 29 | 4.18 | 3.33 | 2.93 | 2.70 | 2.55 | 2.43 | 2.35 | 2.28 | 2.22 |
| 30 | 4.17 | 3.32 | 2.92 | 2.69 | 2.53 | 2.42 | 2.33 | 2.27 | 2.21 |
| 40 | 4.08 | 3.23 | 2.84 | 2.61 | 2.45 | 2.34 | 2.25 | 2.18 | 2.12 |
| 60 | 4.00 | 3.15 | 2.76 | 2.53 | 2.37 | 2.25 | 2.17 | 2.40 | 2.04 |
| 120 | 3.92 | 3.07 | 2.68 | 2.45 | 2.29 | 2.17 | 2.09 | 2.02 | 1.96 |
| ∞ | 3.84 | 3.00 | 2.60 | 2.37 | 2.21 | 2.10 | 2.01 | 1.94 | 1.88 |


*Course Notes by Dr. Hamdi Ghonim - Dr. Hoda Ezzat Kafl* 52